Rate-Distortion-Based Physical Layer Secrecy with Applications to Multimode Fiber

Eva C. Song, *Student Member, IEEE,* Emina Soljanin, *Fellow, IEEE,* Paul Cuff, *Member, IEEE,* H. Vincent Poor, *Fellow, IEEE,* and Kyle Guan, *Member, IEEE*

***Abstract*—Optical networks are vulnerable to physical layer attacks; wiretappers can improperly receive messages intended for legitimate recipients. Multimode fiber (MMF) transmission can be modeled via a broadcast channel in which both the legitimate receiver's and wiretapper's channels are multiple-input-multiple-output complex Gaussian channels. This work considers the theoretical aspects of this security problem in the domain of a broadcast channel. Source-channel coding analyses based on the use of distortion as the metric for secrecy are developed. Alice has a source sequence to be encoded and transmitted over this broadcast channel so that the legitimate user Bob can reliably decode it while forcing the distortion of the wiretapper, or eavesdropper, Eve's estimate to be as high as possible. Tradeoffs between transmission rate and distortion under two extreme scenarios are examined: the best case where Eve has only her channel output and the worst case where she also knows the past realization of the source. It is shown that under the best case, an operationally separate source-channel coding scheme guarantees maximum distortion at the same rate as needed for reliable transmission. Theoretical bounds are given, and particularized for MMF. Numerical results showing the rate distortion tradeoff are presented and compared with corresponding results for the perfect secrecy case.**

***Index Terms*—rate-distortion, MIMO, optical fiber communication, source-channel coding, secrecy, SDM, MMF**

I. Introduction

SINGLE mode fiber systems are believed to have reached their capacity limits. In particular, techniques such as wavelength-division multiplexing (WDM) and polarization-division multiplexing (PDM) have been heavily exploited in the past few years, leaving little room for further improvement in capacity [1]. Space-division multiplexing (SDM) is a promising solution for meeting the growing capacity demands of optical communication networks. One way of realizing SDM is via the use of multimode fiber (MMF). While multimode transmission provides greater capacity, the security of such systems can be an issue because a wiretapper can eavesdrop upon MMF communication by simply bending the fiber [2] . MMF is a multiple-input-multiple-output (MIMO) system [1] that captures the charateristics of crosstalk among different modes. The secrecy capacity of a Gaussian MIMO broadcast channel was studied in [3], but the result cannot be applied directly to MMF because the channel is not the same. The secrecy capacity of this channel was studied in [2] where it is shown that the channel conditions required for perfect secrecy are quite demanding.

The concepts of "perfect secrecy", "partial secrecy", "strong secrecy", "weak secrecy", "equivocation" and "distortion" will be applied repeatedly in this paper. We shall now briefly summarize the relationships among those terms. Please note that, even though "perfect secrecy" is a non-asymptotic concept, here for convenience, we refer to "perfect secrecy" in the asymptotic sense, i.e. the information leakage is arbitrarily small as the blocklength goes to infinity. Information theoretic secrecy typically considers one of the two regimes, perfect secrecy or partial secrecy. Perfect secrecy essentially requires no information leakage to an eavesdropper. In the regime of partial secrecy, one must quantify the degree of secrecy obtained. Equivocation rate is a metric found in the literature that measures how much of the signal is leaked to the eavesdropper regardless of whether the eavesdropper can use the leaked information in a constructive way. Another approach, taken in this work, is to use distortion to measure the difference between the original content and the eavesdropper's estimate of it.

The notions of strong and weak secrecy both refer to the perfect secrecy regime, and equivocation is usually involved in the analysis. Under either strong or weak secrecy, distortion is at the maximum, because negligible information is leaked to the eavesdropper that would allow her to make a better estimate. However, implication in the other direction does not hold in general. In order to keep the distortion at a maximum, neither strong nor weak secrecy is necessarily required.

This work focuses on the partial secrecy regime. It should be noted that equivocation and distortion are not two independent measures. As pointed out in recent work [4], equivocation becomes a special case of distortion when causal information is revealed to the eavesdropper and the distortion is measured by log loss. It can be seen from our results herein that high distortion can be achieved even if all the past information of the source is given to the eavesdropper. Furthermore, partial secrecy comes at a much lower cost than perfect secrecy.

Distortion was also used in [5] and [6] as a metric for secrecy in the context of a noiseless network with secret key sharing. In this work, we are concerned with physical layer secrecy in MMF systems. This prompts us to formulate the problem as a source-channel coding problem along the lines studied in a general setting in [7], some results of which can be directly applied to MMF systems.

Manuscript received April 6, 2013; revised October 1 and December 16, 2013. The editor coordinating the review of this paper and approving it for publication was A. Thangaraj.

This research was supported in part by the National Science Foundation under Grants CCF-1116013 and CNS-09-05086, and in part by the Air Force Office of Scientific Research under Grant FA9550-12-1-0196.

E. C. Song, P. Cuff, and H. V. Poor are with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA (e-mail: {csong, cuff, poor}@princeton.edu).

E. Soljanin and K. Guan are with Bell Labs, Alcatel-Lucent, Murray Hill, NJ 07974, USA (e-mail: {emina, kyle.guan}@alcatel-lucent.com).

Digital Object Identifier 10.1109/TCOMM.2014.010914.130256

The rest of the paper is structured as follows. In Section II, we introduce the system model in two ways: the general source-channel coding model for theoretical derivation; and the particular MMF channel model for application. In Section III, we provide theoretical bounds with source-channel coding for general broadcast channels. This is our main theoretical contribution of the paper. In Section IV, we apply the general results from Section III to the MMF model when the channel is time-invariant and discuss the secrecy outage in the case of a random channel in which the channel state information (CSI) is not available to the transmitter. In Section V, we provide numerical evaluation to the MMF source-channel model under Hamming distortion. Finally, in Section VI, we conclude the paper and discuss open problems from this work.

II. System Model

We first introduce some notation that will be used throughout this paper. A sequence $X_1, ..., X_n$ is denoted by X^n. A limit taken with respect to "$n \to \infty$" is abbreviated as "$\to_n$". In the case that X is a random variable, x is used to denote a realization and $\mathcal{X}$ is used to denote the support of that random variable. $\mathbb{R}$ and $\mathbb{C}$ are reserved to denote the real field and complex field, respectively. A complex Gaussian distribution is denoted by $\mathcal{CN}(\mu, \Sigma, C)$, where μ is the mean, Σ is the covariance matrix, and C is the relation matrix. A Markov relation is denoted by the symbol $-\Box-$. The total variation distance between two distributions P and Q is denoted by $||P - Q||_{TV}$. For a distortion measure $d : \mathcal{S} \times \mathcal{T} \mapsto \mathbb{R}^+$, the distortion between two sequences is defined to be the per-letter average distortion $d(s^k, t^k) = \frac{1}{k}\sum_{i=1}^{k} d(s_i, t_i)$. The maximum distortion Δ, the average distortion achieved by guesses based only on the prior distribution of the source, is defined as

$$\Delta \triangleq \min_{t} \mathbb{E}[d(S,t)]. \tag{1}$$

A. Source-Channel Coding Model for General Broadcast Channels

A source node (Alice) has an independent and identically distributed (i.i.d.) sequence S^k that she intends to transmit over a memoryless broadcast channel $P_{YZ|X}$ such that a legitimate user (Bob) can reliably decode the source sequence, while keeping the distortion between an eavesdropper (Eve) and the source as high as possible. The source sequence S^k is mapped to the channel input sequence X^n through a source-channel encoder. Upon receiving Y^n, Bob makes an estimate $\hat{S}^k$ of the original source sequence S^k. Let $f_{k,n} : \mathcal{S}^k \mapsto \mathcal{X}^n$ be a source-channel encoder and $g_{k,n} : \mathcal{Y}^n \mapsto \mathcal{S}^k$ be the corresponding decoder. For almost lossless reconstruction, we require that the probability that Bob's reconstruction differs from the original goes to zero asymptotically with the source blocklength. That is,

$$\mathbb{P}\left[S^k \neq \hat{S}^k\right] \to_k 0.$$

Similarly, Eve also makes an estimate T^k of S^k upon receiving Z^n and some other side information. We will examine two extreme cases based on the amount of side information Eve has.

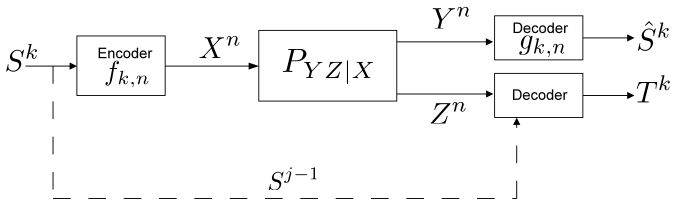


Fig. 1: Source-channel coding model with an i.i.d. source and broadcast channel.

1) No causal information available to Eve: The case in which Eve has only her own channel output but no side information about the source corresponds to the best scenario for the legitimate users of the network, Alice and Bob. With the channel output alone, Eve has very limited resources at hand to make the estimate. Let t^k be Eve's estimate of the original source sequence s^k. The system model is shown in Fig. 1; however, the dashed line represents additional information that is not available to the eavesdropper in this first scenario. We use the lower case $t^k(z^n)$ to denote Eve's deterministic estimation functions of her observation z^n and the capital letter $T^k = t^k(Z^n)$ to denote the function of the random sequence Z^n. The following definitions in this section are for time-invariant channels.

Definition 1. *For a given distortion function $d(s,t)$, a rate distortion pair (R, D) is achievable if there exists a sequence of encoder/decoder pairs $f_{k,n}$ and $g_{k,n}$ such that*

$$\frac{k}{n} = R,$$

$$\lim_{n\to\infty} \mathbb{P}\left[S^k \neq \hat{S}^k\right] = 0,$$

and

$$\liminf_{n\to\infty} \min_{t^k(z^n)} \mathbb{E}\left[d(S^k, t^k(Z^n))\right] \geq D.$$

Note that the rate-distortion pair (R, D) captures the tradeoff between Bob's rate for reliable transmission and Eve's distortion, which is different from rate-distortion theory in the traditional sense.

2) With causal information available to Eve: On the other hand, we are also interested in the case in which, at each time instance j, Eve gets to see the past realization of the source sequence S^{j-1}. This would be the worst scenario for the legitimate users. The definition for an achievable rate distortion pair (R, D) is similar to Definition 1 given in the previous subsection except the last condition is replaced by

$$\liminf_{n\to\infty} \min_{\{t_j(z^n, s^{j-1})\}_{j=1}^{k}} \mathbb{E}\left[\frac{1}{k}\sum_{j=1}^{k} d(S_j, t_j(Z^n, S^{j-1})\right] \geq D.$$

The system model is shown in Fig. 1 with the dashed line representing the availability of the causal information.

B. MMF Channel Model

Now we particularize the general broadcast channel described above to an MMF broadcast channel as shown in Fig. 2. An M-mode MMF is modeled as a memoryless MIMO channel with input X an M-dimensional complex vector. Here M is a positive integer.

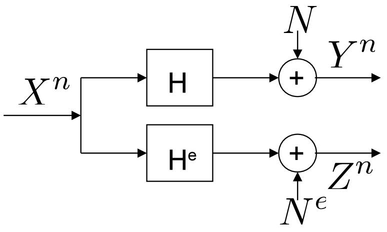


Fig. 2: MMF channel model.

Unlike wireless MIMO which has a total power constraint, MMF channels have the following per mode power constraint averaged over n uses of the channel:

$$\frac{1}{n}\sum_{i=1}^{n}\left|X_i^{(m)}\right|^2 \leq 1 \quad \text{for all modes } m \in [1:M]. \tag{2}$$

More generally (as in [3]), we will consider a power constraint of the form

$$\frac{1}{n}\sum_{i=1}^{n} X_i X_i^{\dagger} \quad \preceq \quad Q, \tag{3}$$

where $Q \in \{A \in \mathcal{H}^{M\times M} : A \succeq 0, A_{ii} = 1\}$ and $\mathcal{H}$ denotes the set of Hermitian matrices. One element in this set is the identity matrix I (constraint (2)). We will focus on the case $Q = I$ for simplicity. A detailed discussion of the MMF channel model can be found in [1].

1) The Legitimate User Communications Model: The channel between Alice and Bob $P_{Y|X}$ is complex, Gaussian, MIMO, with input $X \in \mathbb{C}^M$ as described above, and output $Y \in \mathbb{C}^M$ given by

$$Y \quad = \quad HX + N, \tag{4}$$

where $N \sim \mathcal{CN}(0, \sigma_N^2 I, 0)$ is M-dimensional, uncorrelated, zero-mean, complex, Gaussian noise and H is an $M \times M$ complex matrix. Bob's channel matrix H is of the form

$$H \quad = \quad \sqrt{E_0 L}\Psi, \tag{5}$$

where $\Psi \in \mathbb{C}^{M\times M}$ is unitary and $E_0 L$ is a constant scalar that measures the average power of the channel. We refer to $E_0 L/\sigma_N^2$ as the SNR of the channel. Matrix Ψ, the unitary factor of the channel H, describes the modal crosstalk [1].

2) The Eavesdropper Communications Model: The channel between Alice and Eve $P_{Z|X}$ is also complex, Gaussian, MIMO, with input $X \in \mathbb{C}^M$ as described above, and output $Z \in \mathbb{C}^M$ given by

$$Z \quad = \quad H^e X + N^e, \tag{6}$$

where $N^e \sim \mathcal{CN}(0, \sigma_{N^e}^2 I, 0)$ is M-dimensional uncorrelated, zero-mean, complex, Gaussian noise, and H^e is an $M \times M$ complex matrix. Eve's channel matrix H^e is of the form

$$H^e \quad = \quad \sqrt{E_0 L^e}\sqrt{\Phi}\Psi^e, \tag{7}$$

where $\Psi^e \in \mathbb{C}^{M\times M}$ is unitary, Φ is diagonal with positive entries, and $E_0 L^e$ is the average power of Eve's channel. Note that Eve has a different signal to noise ratio $\text{SNR}^e = E_0 L^e/\sigma_{N^e}^2$. The diagonal component Φ of the channel matrix H^e corresponds to the mode-dependent loss (MDL) as introduced in [1].

III. Theoretical Bounds

In this section, we focus on the general broadcast channel introduced in Section II-A only. We first make some general observations about the communication between Alice and Bob, as well as the communication between Alice and Eve. If Eve is not present, Alice and Bob can communicate losslessly at any rate lower than $R_0 \triangleq \frac{\max_X I(X;Y)}{H(S)}$ because separate source-channel coding is optimal for point-to-point communication. Ideally, we want to force maximum distortion Δ upon Eve. But higher distortion to Eve may come at the price of a lower communication rate to Bob. The technical content of this section is organized as follows: the rate-distortion region for the "no causal information" case is first given in Theorem 1; to prepare for the achievability proof of Theorem 1, an operational separation scheme is discussed; also, an achievable rate-distortion region is given in Theorem 5 for the causal case under Hamming distortion; and finally, an example with a binary symmetric channel and Hamming distortion is provided for illustration.

Before starting the new results, we shall provide a recap of what has been done in the literature regarding this problem and what our main advances are in this work. For noiseless channels, the source coding problems of both the no-causal-information and the causal-information cases were solved in [8] and [6], respectively. There, secrecy was obtained by using a secret key shared between Alice and Bob. As for physical layer secrecy of a memoryless broadcast channel, the result for transmitting two messages, one confidential and one public, from Csiszár and Körner [9] have been known for many decades. In their work, weak secrecy was considered. This result was strengthened in [10] by considering strong secrecy. The same rate region was obtained in [10], however the metric for secrecy is stronger. In our work, the source-channel coding schemes we propose operationally separate source and channel coding that requires dividing the bit sequence produced by source coding into two messages which are then processed by the channel coding. The channel coding part functions in a way that is similar to [9] or [10], except that the public message in those work is not required to be decoded in our case, and we refer to that message as the "non-confidential" message. This type of channel coding setting was also used in [7] for the causal-information case and it is shown there that only weak secrecy is required to combine the source and channel coding. In this work, we will connect the source coding (from [8]) and channel coding for the no-causal-information case. Unlike the causal-information case, strong secrecy from the channel is needed. This will require modifying some of the settings from [10]. We also extend the result for the causal-information case from [7] to include the rate-distortion tradeoff.

We now state the rate-distortion result for general source-channel coding with an i.i.d. source sequence and a discrete memoryless broadcast channel $P_{YZ|X}$ when **no causal information** is available to Eve. In the following theorem, we will see that the source sequence can be delivered almost losslessly to Bob at a rate arbitrily close to R_0 while the distortion to Eve is kept at Δ, as long as the secrecy capacity is positive.

Theorem 1. *For an i.i.d. source sequence S^k and memoryless broadcast channel $P_{YZ|X}$, if there exists W−□−X−□−YZ*

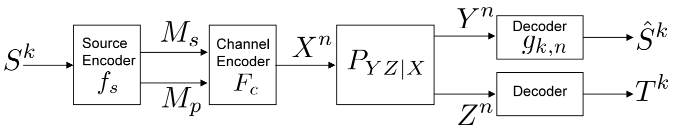


Fig. 3: Operationally separate source-channel coding: the confidential and non-confidential messages satisfy $M_s \in [1 : 2^{kR'_s} = 2^{nR_s}]$ and $M_p \in [1 : 2^{kR'_p} = 2^{nR_p}]$.

such that $I(W;Y) - I(W;Z) > 0$, then (R, D) is achievable if and only if

$$R < \frac{\max_X I(X;Y)}{H(S)}, \tag{8}$$

$$D \leq \Delta, \tag{9}$$

where Δ was defined in (1).

Remark: The requirement $I(W;Y) - I(W;Z) > 0$ implies the existence of a secure channel with a positive rate, i.e. the eavesdropper's channel is not less noisy than the intended receiver's channel. So instead of demanding a high secure transmission rate with perfect secrecy to accommodate the description of the source, we need only to ensure the existence of a secure channel with positive rate. This will suffice to ensure that the eavesdropper's distortion is maximal.

The converse is straightforward. Each of the inequalities (8) and (9) is true individually for any channel and source, (8) by channel capacity coupled by optimality of source-channel separation, and (9) by definition.

The idea for achievability is to operationally separate the source and channel coding (see Fig. 3). The source encoder compresses the source and splits the resulting message into a confidential message and a non-confidential message. A channel encoder is concatenated digitally with the source encoder so that the channel delivers both the confidential and non-confidential messages reliably to Bob and keeps the confidential message secret from Eve, as in [9]. The overall source-channel coding rate will have the following form: $R = \frac{k}{n} = \frac{k}{\log|\mathcal{M}|} \cdot \frac{\log|\mathcal{M}|}{n} = \frac{R_{ch}}{R_{src}}$, where $|\mathcal{M}|$ is the total cardinality of the confidential and the non-confidential messages; R_{ch} and R_{src} are the channel coding and source coding rates, respectively.

Let us look at two models in the following subsections that will help us establish the platform for showing the achievability of Theorem 1.

A. Channel Coding and Strong Secrecy

Consider a memoryless broadcast channel $P_{YZ|X}$ and a communication system with a confidential message M_s and a non-confidential message M_p that must allow the intended receiver to decode both M_s and M_p while keeping the eavesdropper from learning anything about M_s. Problems like this were first studied by Csiszár and Körner [9] in 1978, as an extension of Wyner's work in [11]. However, their model and our model differ in that the second receiver in their setting is required to decode the public message M_p. The mathematical formulation and result of our channel model is stated below. We focus on the message pairs (M_s, M_p) whose distribution satisfies the following:

$$P_{M_s|M_p=m_p}(m_s) \quad = \quad 2^{-nR_s} \tag{10}$$

for all (m_s, m_p). Later we will show that a source encoder can always prepare the input messages to the channel in this form.

Definition 2. *An (R_s, R_p, n) channel code consists of a channel encoder F_c (possibly stochastic) and a channel decoder g_c such that*

$$F_c : \mathcal{M}_s \times \mathcal{M}_p \mapsto \mathcal{X}^n$$

and

$$g_c : \mathcal{Y}^n \mapsto \mathcal{M}_s \times \mathcal{M}_p$$

where $|\mathcal{M}_s| = 2^{nR_s}$ and $|\mathcal{M}_p| = 2^{nR_p}$.

Definition 3. *The rate pair (R_s, R_p) is achievable under weak secrecy if for all (M_s, M_p) satisfying* (10)*, there exists a sequence of (R_s, R_p, n) channel codes such that*

$$\lim_{n\to\infty} \mathbb{P}\left[(M_s, M_p) \neq (\hat{M}_s, \hat{M}_p)\right] = 0$$

and

$$\lim_{n\to\infty} \frac{1}{n} I(M_s; Z^n | M_p) = 0.$$

Note that because the eavesdropper may completely or partially decode M_p, the secrecy requirement is modified accordingly to consider $I(M_s; Z^n|M_p)$ instead of $I(M_s; Z^n)$. To guarantee true secrecy of M_s, we want to make sure that even if M_p is given to the eavesdropper, there is no information leakage of M_s, because $I(M_s; Z^n|M_p) = I(M_s; Z^n, M_p)$ if M_s and M_p are independent.

Theorem 2 (Theorem 3 in [7])**.** *A rate pair (R_s, R_p) is achievable under weak secrecy if*

$$R_s \quad \leq \quad I(W;Y|V) - I(W;Z|V), \tag{11}$$

$$R_p \quad \leq \quad I(V;Y) \tag{12}$$

for some V –□– W –□– X –□– YZ.

The proof can be found in [7]. Let us denote the above region as $\mathcal{R}$. We now strengthen the result by considering strong secrecy introduced in [12]. Later we will use strong secrecy to connect the operationally separate source and channel encoders.

Definition 4. *The rate pair (R_s, R_p) is achievable under strong secrecy if for all (M_s, M_p) satisfying* (10)*, there exists a sequence of (R_s, R_p, n) channel codes such that*

$$\lim_{n\to\infty} \mathbb{P}[(M_p, M_s) \neq (\hat{M}_s, \hat{M}_p)] = 0$$

and

$$\lim_{n\to\infty} I(M_s; Z^n|M_p) = 0.$$

In general, weak secrecy does not necessarily imply that strong secrecy is also achievable; however, in this particular setting we have the following claim:

Theorem 3. *A rate pair (R_s, R_p) achievable under weak secrecy is also achievable under strong secrecy.*

The following two lemmas will assist the proof of Theorem 3 by providing a sufficient condition for satisfying the secrecy constraint $\lim_{n\to\infty} I(M_s; Z^n|M_p) = 0$.

Lemma 1. *If*

$$||P_{Z^n|M_p=m_p}P_{M_s|M_p=m_p} - P_{Z^nM_s|M_p=m_p}||_{TV} \leq \epsilon \leq \frac{1}{2},$$

then

$$I(M_s;Z^n|M_p=m_p) \leq -\epsilon \log \frac{\epsilon}{|\mathcal{M}_s|}.$$

Proof: Let $\epsilon_{z^n} = \left|\left|P_{M_s|M_p=m_p} - P_{M_s|Z^n=z^n,M_p=m_p}\right|\right|_{TV}$. Therefore,

$$\begin{aligned}&\mathbb{E}_{P_{Z^n|M_p=m_p}}[\epsilon_{z^n}]\\ = \;& \left|\left|P_{Z^n|M_p=m_p}P_{M_s|M_p=m_p} - P_{Z^nM_s|M_p=m_p}\right|\right|_{TV} \leq \epsilon.\end{aligned}$$

By Lemma 2.7 [13],

$$\begin{aligned}&|H(M_s|M_p=m_p) - H(M_s|Z^n=z^n,M_p=m_p)|\\ \leq \;& -\epsilon_{z^n}\log\frac{\epsilon_{z^n}}{|\mathcal{M}_s|}.\end{aligned}$$

Note that $f(x) \triangleq -x\log x$ is concave. And by applying Jensen's inequality twice, we have

$$\begin{aligned}&I(M_s;Z^n|M_p=m_p)\\ = \;& |\mathbb{E}_{P_{Z^n|M_p=m_p}}[H(M_s|M_p=m_p)\\ &\qquad -H(M_s|Z^n=z^n,M_p=m_p)]|\\ \leq \;& \mathbb{E}_{P_{Z^n|M_p=m_p}}[|H(M_s|M_p=m_p)\\ &\qquad -H(M_s|Z^n=z^n,M_p=m_p)|]\\ \leq \;& \mathbb{E}_{P_{Z^n|M_p=m_p}}\left[-\epsilon_{z^n}\log\frac{\epsilon_{z^n}}{|\mathcal{M}_s|}\right]\\ \leq \;& -\epsilon\log\frac{\epsilon}{|\mathcal{M}_s|}.\end{aligned}$$

■

Lemma 2. *If for every (m_s, m_p), there exists a measure θ_{m_p} on $\mathcal{Z}^n$ such that*

$$||P_{Z^n|M_p=m_p,M_s=m_s} - \theta_{m_p}||_{TV} \leq \epsilon_n$$

then

$$\lim_{n\to\infty} I(M_s;Z^n|M_p) = 0$$

where $\epsilon_n = 2^{-n\beta}$ for some $\beta > 0$.

A proof of Lemma 2 is given in Appendix A.

If there exist channel codes such that $\mathbb{P}\left[(M_s,M_p) \neq (\hat{M}_s,\hat{M}_p)\right] \to_n 0$ and measure θ_{m_p} for all (m_s,m_p) such that $||P_{Z^n|M_p=m_p,M_s=m_s} - \theta_{m_p}||_{TV} \leq \epsilon_n$, then Theorem 3 follows immediately. The existence of such a code and measure is assured by the same codebook construction and choice of measure as in [10].

B. Source Coding

Recall from our problem formulation in Section II that the sender Alice has an i.i.d. source sequence S^k. A source encoder is needed to prepare S^k by encoding it into a pair of messages (M_s, M_p) that satisfies $P_{M_s|M_p=m_p}(m_s) = 2^{-kR'_s} = 2^{-nR_s}$ so that it forms a legitmate input to the channel model in Section III-A.

Definition 5. *An (R'_s, R'_p, k) source code consists of an encoder f_s and a decoder g_s such that*

$$f_s : \mathcal{S}^k \mapsto \mathcal{M}_s \times \mathcal{M}_p$$

$$g_s : \mathcal{M}_s \times \mathcal{M}_p \mapsto \mathcal{S}^k$$

where $|\mathcal{M}_s| = 2^{kR'_s}$ and $|\mathcal{M}_p| = 2^{kR'_p}$.

Definition 6. *A rate distortion triple (R'_s, R'_p, D) is achievable under a given distortion measure $d(s,t)$ if there exists a sequence of (R'_s, R'_p, k) source codes such that*

$$\lim_{k\to\infty}\mathbb{P}\left[S^k \neq g_s(f_s(S^k))\right] = 0$$

and the message pair generated by the source encoder satisfies $P_{M_s|M_p=m_p}(m_s) = 2^{-kR'_s}$ and for all $P_{Z^n|M_sM_p}$ such that $I(M_s;Z^n|M_p) \to_n 0$

$$\liminf_{k\to\infty}\min_{t^k(z^n)}\mathbb{E}\left[d^k(S^k,t^k(Z^n))\right] \geq D.$$

Theorem 4. *(R'_s, R'_p, D) is achievable if*

$$R'_s > 0,$$

$$R'_s + R'_p > H(S),$$

and

$$D \leq \Delta.$$

The general idea for achievability is to consider the ϵ-typical S^k sequences and partition them into bins of equal size so that each bin contains sequences of the same type. The identity M_p of the bin is revealed to all parties, but the identity M_s of each sequence inside a bin is perfectly protected.[1] Each such partition is treated as a codebook. It was shown in [8] that, for the noiseless case in which Eve is given m_p instead of z^n, the distortion averaged over all such codebooks achieves the maximum distortion Δ as $k \to \infty$ and therefore there must exist one partition that achieves Δ. In order to transition from the result in [8] to our claim in Theorem 4, we only need to show

$$\min_{t^k(z^n)}\mathbb{E}\left[d^k(S^k,t^k(Z^n))\right] \geq \min_{t^k(m_p)}\mathbb{E}\left[d^k(S^k,t^k(M_p))\right].$$

Proof: First, observe that

$$\begin{aligned}&\min_{t^k(\cdot)}\mathbb{E}\left[d^k(S^k,t^k(\cdot))\right]\\ = \;& \frac{1}{k}\sum_{i=1}^{k}\min_{t(i,\cdot)}\mathbb{E}\left[d(S_i,t(i,\cdot))\right]. \qquad (13)\end{aligned}$$

Next, we claim the channel output sequence z^n does not provide Eve anything more than m_p and therefore

$$\begin{aligned}&\min_{t(i,z^n)}\mathbb{E}\left[\frac{1}{k}\sum_{i=1}^{k}d(S_i,t(i,Z^n))\right]\\ \geq \;& \min_{t(i,m_p)}\mathbb{E}\left[\frac{1}{k}\sum_{i=1}^{k}d(S_i,t(i,M_p))\right] - 2\delta'(\epsilon). \qquad (14)\end{aligned}$$

The analysis is similar to that in [7], but for the sake of clarity, we present the complete proof of (14) in Appendix B. Here

[1] Strictly speaking, the source encoder may violate the condition (10) on $(k+1)^{|\mathcal{S}|}$ number of bins, because $(k+1)^{|\mathcal{S}|}$ is an upper bound on the number of types of a sequence of length k. However, this is just a very small (polynomial in k) number of bins compared with the total number (roughly $2^{kH(S)}$) of bins. Therefore, for this small portion of "bad" bins that violate (10), we can just let the source encoder declare an error on the confidential message M_s and construct a dummy M_s uniformly given the bin index m_p. This will contribute only an ϵ factor to the error probability.

strong secrecy comes into play. It is also pointed out within the proof in Appendix B that $I(M_s; Z^n|M_p) \to_n 0$ is needed.

Finally, combining (14) with (13) give us the desired result. ■

C. Achievability of Theorem 1

With all the elements from Section III-A and III-B, we are now ready to complete the achievability proof of Theorem 1 using Theorems 2 and 4 by concatenating the channel encoder with the source encoder.

Proof: Fix $\nu \geq \epsilon > 0$. Fix P_S. Let ${R_s}' = 2\nu$, ${R_p}' = H(S) - \nu$ and $R' = {R_s}' + {R_p}'$. We apply the same codebook construction and encoding scheme as in Section III-B by partioning the ϵ-typical S^k sequences into $2^{k{R_p}'}$ bins and inside each bin we have $2^{k{R_s}'}$ sequences so that $\mathbb{P}[S^k \neq g_s(f_s(S^k))] \leq \epsilon$. Recall that all the sequences inside one bin are of the same type, so it is guaranteed that

$$P_{M_s|M_p=m_p}(m_s) = \frac{1}{|\mathcal{M}_s|} = \frac{1}{2^{kR'_s}}$$

for all m_p, m_s, which implies $I(M_s; M_p) = 0$.

Let R_s and R_p be the channel rates. R_p is seen as a function of R_s on the boundary of the region given in Theorem 2 and this is denoted by $R_p(R_s)$. Suppose $\max_{(R_s,R_p)\in\mathcal{R}} R_s > 0$, i.e. there exists $W - \square - X - \square - YZ$ such that $I(W; Y) - I(W; Z) > 0$ (justified in Appendix C). $R_p(R_s)$ is continuous and non-increasing. Thus, R_p achieves its maximum at $R_s = 0$, which would be the channel capacity $\max_X I(X; Y)$ of $P_{Y|X}$ for reliable transmission. By the continuity of $R_p(R_s)$, $(R_s, R_p) = (2\nu\frac{k}{n}, R_p(0) - \delta(\nu))$ is achievable under strong secrecy, i.e. $\mathbb{P}[(M_s, M_p) \neq (\hat{M}_s, \hat{M}_p)] \leq \epsilon$ and $I(M_s; Z^n|M_p) \leq \epsilon$, where $\delta(\nu) \to 0$ as $\nu \to 0$.

From the above good channel code under strong secrecy we have $P_{Z^n|M_sM_p}$ such that $I(M_s; Z^n|M_p) \to_n 0$. Therefore, we can apply Theorem 4 to achieve

$$\liminf_{k\to\infty} \min_{t^k(z^n)} \mathbb{E}\left[d^k(S^k, t^k(Z^n))\right] = D.$$

The error probability is bounded by the sum of the error probabilities from the source coding and channel coding parts, i.e. $\mathbb{P}\left[S^k \neq \hat{S}^k\right] < 2\epsilon$. Finally, we verify the total transmission rate to complete the proof:

$$\begin{aligned} R &= \frac{k}{n} = \frac{R_s + R_p}{R'_s + R'_p} \\ &= \frac{R_p(0) - \delta(\nu) + 2R\nu}{H(S) + \nu} \\ &\geq \frac{R_p(0) - \delta(\nu)}{H(S) + \nu} \\ &\overset{\nu\to 0}{\longrightarrow} \frac{\max_X I(X; Y)}{H(S)}. \end{aligned}$$

■

We next state the rate-distortion result for source-channel coding with an i.i.d. source sequence and discrete memoryless broadcast channel $P_{YZ|X}$ when **causal information** is available to Eve. The result comes from the rate matching of [7].

Theorem 5. *For an i.i.d. source sequence S^k and a memoryless broadcast channel $P_{YZ|X}$, a rate distortion pair (R, D) is achievable if*

$$\begin{aligned} R &\leq \min\left(\frac{I(V;Y)}{I(S;U)}, \frac{I(W;Y|V) - I(W;Z|V)}{H(S|U)}\right), \\ D &\leq \frac{\alpha}{R} \cdot \Delta + \left(1 - \frac{\alpha}{R}\right) \cdot \min_{t(u)} \mathbb{E}\left[d(S, t(U))\right] \end{aligned}$$

for some distribution $P_S P_{U|S} P_V P_{W|V} P_{X|W} P_{YZ|X}$, where $\alpha = \frac{[I(V;Y) - I(V;Z)]^+}{I(S;U)}$.

Example: binary symmetric broadcast channel (BSBCC) and binary source with Hamming distortion

To visualize Theorem 1 and Theorem 5, we will illustrate the results with a BSBCC and binary source under Hamming distortion, defined as

$$d_H(s,t) = \begin{cases} 0, & s = t, \\ 1, & \text{otherwise.} \end{cases}$$

With the above setting, suppose $S_i \sim \text{Bern}(p)$, and the broadcast channel is binary symmetric with crossover probabilities to the intended receiver and the eavesdropper p_1 and p_2, respectively. Assume $p \leq 0.5$ and $p_1 < p_2 < 0.5$. It is well known that this can be treated as a physically degraded channel in capacity calculation. Let us make the following definitions:

$$f(x) \text{ is the linear interpolation of the points} \\ \left(\log n, \frac{n-1}{n}\right), n = 1, 2, 3, ... \tag{15}$$

$$d(x) \triangleq \min(f(x), 1 - \max_s P_S(s)), \tag{16}$$

$$h(x) \triangleq x \log\frac{1}{x} + (1-x)\log\frac{1}{1-x} \\ \text{is the binary entropy function,} \tag{17}$$

$$x_1 * x_2 \triangleq x_1 * (1 - x_2) + (1 - x_1) * x_2 \\ \text{is the binary convolution,} \tag{18}$$

where $P_S(\cdot)$ is the probability mass function of the random variable S. The corresponding rate-distortion regions for the no-causal-information and causal-information cases are given in the following corollaries.

Corollary 1. *For an i.i.d. Bern(p) source sequence S^k and BSBCC with crossover probabilities p_1 and p_2, when **no causal information** is available, (R, D) is achievable if and only if*

$$\begin{aligned} R &< \frac{1 - h(p_1)}{h(p)}, \\ D &\leq p. \end{aligned}$$

Corollary 2. *For an i.i.d. Bern(p) source sequence S^k and BSBCC with crossover probabilities p_1 and p_2, when **causal information** is available, (R, D) is achievable if*

$$\begin{aligned} R &\leq \frac{h(p_2) - h(p_1)}{h(p)}, \\ D &\leq p \end{aligned}$$

or

$$\frac{h(p_2) - h(p_1)}{h(p)} < R \leq \frac{1 - h(p_1)}{h(p)},$$

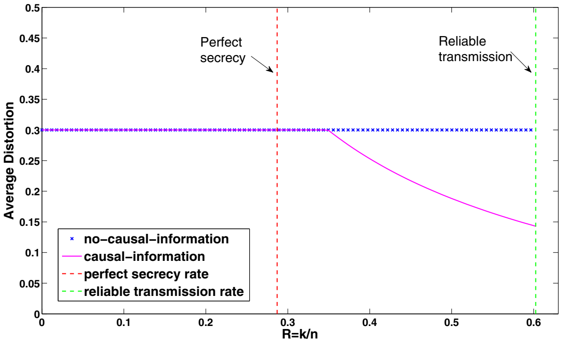


Fig. 4: Achievable distortion-rate curves. On the horizontal axis is the symbol/channel use source-channel coding rate and on the vertical axis is the average Hamming distortion.

$$D \leq \alpha' p + (1-\alpha') d\left(\frac{h(\gamma * p_1) - h(\gamma * p_2) - h(p_1) + h(p_2)}{R}\right)$$

where $\gamma \in [0, 0.5]$ *solves* $h(\gamma * p_2) = 1 - h(p_1) + h(p_2) - Rh(p)$ *and* $\alpha' = \frac{h(\gamma * p_2) - h(\gamma * p_1)}{1 - h(\gamma * p_1)}$.

These corollaries result directly from applying Theorem 1 and Theorem 5, respectively. The region given in Corollary 2 is calculated in a similar fashion as the region given by Theorem 7 of [7]. A numerical example with $p = 0.3$, $p_1 = 0.1$ and $p_2 = 0.2$ is plotted in Fig.4. Interpretation of the plot is deferred until the end of Section V.

IV. MMF Main Results

A. Fixed MMF Channel

We now apply the above result to the MMF model introduced in Section II-B by finding the rate distortion regions for the MMF model defined in (4) and (6) under the two scenarios. In this section, as before, we assume the channels are time-invariant. First of all, we will give the achievable rate region under strong secrecy (therefore also under weak secrecy).

Theorem 6. *The following rate region for one confidential and one non-confidential message is achievable under strong secrecy for a complex Gaussian channel:*

$$R_s \leq \log \frac{|HKH^\dagger + \sigma_N^2 I|}{|\sigma_N^2 I|} - \log \frac{|H^e K H^{e\dagger} + \sigma_{N^e}^2 I|}{|\sigma_{N^e}^2 I|} \quad (19)$$

$$R_p \leq \log \frac{|HQH^\dagger + \sigma_N^2 I|}{|HKH^\dagger + \sigma_N^2 I|} \quad (20)$$

for some K *and* Q*, where* $0 \preceq K \preceq Q$*,* $K \in \mathcal{H}^{M\times M}$*,* Q *satisfies the power constraint in* (3)*, and* H *and* H^e *are the channel gain matrices.*

Proof: According to Theorem 2 and 3,

$$R_s \leq I(W;Y|V) - I(W;Z|V) \quad (21)$$

$$R_p \leq I(V;Y) \quad (22)$$

for some $V -\!\!\square\!\!- W -\!\!\square\!\!- X -\!\!\square\!\!- YZ$ and $\mathbb{E}[XX^\dagger] \preceq Q$, is an achievable rate pair.

We restrict the channel input X to be a circularly symmetric complex Gaussian vector. Let $V \sim \mathcal{CN}(0, Q-K, 0)$ and $B \sim \mathcal{CN}(0, K, 0)$ such that B and V are independent, and $W = X = V + B$. Therefore, $X \sim \mathcal{CN}(0, Q, 0)$ satisfies the power constraint. Similar to results in [3], the rate pair (R_s, R_p) satisfying inequalities (19) and (20) can be achieved. ■

An immediate corollary follows directly from the above theorem.

Corollary 3. *The following rate pairs are achievable under strong secrecy for MMF with channel gains defined in* (5) *and* (7) *and equal full power allocation* $Q = I$*:*

$$R_s \leq \log \frac{|SNR K + I|}{|SNR^e \Psi^e K \Psi^{e\dagger} \Phi + I|} \quad (23)$$

$$R_p \leq \log \frac{|(SNR + 1)I|}{|SNR K + I|} \quad (24)$$

for some K *where* $0 \preceq K \preceq I$*,* $K \in \mathcal{H}^{M\times M}$*,* $SNR = E_0 L/\sigma_N^2$ *and* $SNR^e = E_0 L^e/\sigma_{N^e}^2$*.*

With the secrecy capacity region of MMF, we can evaluate its rate distortion region (R, D) under the two extreme cases, without and with causal information at Eve's decoder respectively. For the best case scenario (no causal information), we will give a sufficient condition to force maximum distortion Δ between Alice and Eve. For the worst case scenario (with causal information), we will give an achievable rate-distortion region and look at the particular case of Hamming distortion.

Theorem 7. *For an i.i.d source sequence* S^k*, if*

$$\min_{j\in\{1,\ldots,M\}} \bar{\phi}_j < \frac{SNR}{SNR^e} \quad (25)$$

where $\bar{\phi}_j$*'s are the diagonal entries of* Φ*, then the following rate distortion pair* (R, D) *is achievable with* ***no causal information*** *at the eavesdropper:*

$$R < \frac{M \log(SNR + 1)}{H(S)} \quad (26)$$

$$D \leq \Delta. \quad (27)$$

Theorem 7 follows from Theorem 1 and Corollary 3. Note that (25) is a sufficient condition for the existence of a secure channel with strictly positive rate from Alice to Bob. A discussion of this condition is provided in Appendix D.

Theorem 8. *For an i.i.d. source sequence* S^k *and Hamming distortion, the following distortion rate curve* $D(R)$ *is in the achievable region* ***with causal information*** *at the eavesdropper:*

$$D = d(H(S)), \text{ if } R \leq \frac{R_s^*}{H(S)} \quad (28)$$

$$D = \bar{\alpha}(K)\Delta + (1 - \bar{\alpha}(K))\, d\left(\frac{R_s(K)}{R}\right), \quad \text{if } \frac{R_s^*}{H(S)} < R \leq \frac{R_p^*}{H(S)} \quad (29)$$

where $d(\cdot)$ *is as defined in* (16)*;* $\mathcal{K} \triangleq \{K \in \mathcal{H}^{M\times M}, 0 \preceq K \preceq I\}$*,*

$$R_s^* = \max_{K'\in\mathcal{K}} \log \frac{|SNR K' + I|}{|SNR^e \sqrt{\Phi} \Psi^e K' \Psi^{e\dagger} \sqrt{\Phi} + I|},$$

$$R_p^* = M \log(SNR + 1),$$

$$R_s(K) = \log \frac{|SNRK + I|}{|SNR^e \sqrt{\Phi} \Psi^e K \Psi^{e\dagger} \sqrt{\Phi} + I|},$$

$$\bar{\alpha}(K) = \frac{\bar{\beta}(K) - \bar{\gamma}(K)}{\bar{\beta}(K)},$$

$$\bar{\beta}(K) = \log \frac{|(SNR + 1)I|}{|SNRK + I|},$$

$$\bar{\gamma}(K) = \log \frac{|SNR^e \Phi + I|}{|SNR^e \sqrt{\Phi} \Psi^e K \Psi^{e\dagger} \sqrt{\Phi} + I|}.$$

The result given in Theorem 8 can be derived directly from Theorem 5 and Corollary 3.

B. Secrecy Outage under a Random MMF Channel

All the results we have seen thus far were derived for a time-invariant channel, which means that both the transmitter and the receivers are informed about the channel state. However, in MMF, the channels H and H^e vary with time and the CSI is not available at the transmitter due to the long round-trip delay over the large distances common in optical transmission, even though it has a long coherence time, i.e. H and H^e are essentially constant over n channel uses. In Corollary 2, Theorem 7 and Theorem 8, we have chosen $Q = I$ as the channel input power for simplicity. This power allocation strategy also tends to minimize the outage probability for perfect secrecy [14].

The randomness of $H = \sqrt{E_0 L}\Psi$ and $H^e = \sqrt{E_0 L^e}\sqrt{\Phi}\Psi^e$ comes from the unitary component Ψ, Ψ^e and the diagonal component Φ. The random matrices Ψ and Ψ^e are uniformly distributed in $\ominus$, where $\ominus$ is the set of all $M \times M$ unitary matrices [1]. The diagonal matrix $\Phi = diag\{\bar{\phi}_1, ..., \bar{\phi}_M\}$, where $\bar{\phi}_i = M\frac{\phi_i}{\sum_{j=1}^{M}\phi_j}$. Here $\phi_1 = \phi_{min}$ and ϕ_{max}, and $\phi_i \sim Unif[\phi_{min}, \phi_{max}]$ for $i = 3, ..., M$.

In this situation of no CSI at the transmitter with long coherence time, performance is typically measured by outage probability. The capacity $C = \max_Q \log|I + HQH^\dagger|$ and secrecy capacity $C_s = \max_Q \left[\log|I + HQH^\dagger| - \log|I + H^e Q H^{e\dagger}|\right]$ for a deterministic MIMO Gaussian broadcast channel were given in [15] and [16], respectively. For the case of no causal information at the eavesdropper, the CSI does not really affect the performance much. The channel capacity between Alice and Bob $M\log(1 + \text{SNR})$ does not depend on the channel realization due to the unitary component of the channel. Hence, the encoder can choose the source-channel coding rate to be just below $\frac{M\log(\text{SNR}+1)}{H(S)}$, and maximum distortion can be achieved if the channel satisfies the condition in Theorem 7. For the case where Eve has causal source information, we consider only the input power $Q = I$ and Hamming distortion. Without knowledge of CSI, the source-channel encoder chooses a source-channel coding rate $\bar{R}$, a pair of source coding rates $(\bar{R}'_s, \bar{R}'_p)$ on the line segment $R'_s + R'_p = H(S)$, $R'_s, R'_p \geq 0$ and a real value $\alpha \in [0, 1]$. Let

$$\bar{R}_s \triangleq \bar{R}'_s\bar{R} \text{ and } \bar{R}_p \triangleq \bar{R}'_p\bar{R}.$$

We define the outage probability of such a choice of parameters to be

$$\begin{aligned} & P_{out}(I, \bar{R}'_s, \bar{R}, \alpha) \\ = \ & 1 - \sum_{K \in \mathcal{K}} \mathbb{P}_{\Phi\Psi^e\Psi}\left[(\bar{R}_s, \bar{R}_p) \in \mathcal{R}_{\Phi\Psi^e\Psi}(I) \text{ and } \bar{\alpha}(K) \geq \alpha\right], \end{aligned}$$

where $\mathcal{R}_{\Phi\Psi^e\Psi}(I)$ denotes the region given in Corollary 3, $\mathcal{K} = \{K : 0 \preceq K \preceq I, \bar{R}_s = \log\frac{|\text{SNR}K+I|}{|\text{SNR}^e\Psi^e K \Psi^{e\dagger}\Phi+I|}$ and $\bar{R}_p = \log\frac{|(\text{SNR}+1)I|}{|\text{SNR}K+I|}\}$,and $\bar{\alpha}(K)$ is defined as in Theorem 8. Note that $\mathcal{R}_{\Phi\Psi^e\Psi}(Q)$ is a random variable because the channels $P_{YZ|X}$ are now random. Under this set of parameters $(\bar{R}'_s, \bar{R}, \alpha)$, we can achieve distortion $\alpha\Delta + (1-\alpha)d(\bar{R}'_s)$ with probability $1 - P_{out}(I, \bar{R}'_s, \bar{R}, \alpha)$, where $d(\cdot)$ was defined in (16). Proving the existence of a good codebook in this case is an important information theoretic problem, most recently addressed in [17]. For some recent progress, please refer to [17].

V. Numerical Results

In this section, we present numerical results illustrating achievable rate distortion regions of an MMF under the two information models with a time-invariant channel. Let us consider measuring the eavesdropper's distortion using Hamming distortion and a Bern(p) i.i.d. source sequence. Fig. 3 shows numerical results corresponding to Theorem 7 and Theorem 8 under equal power allocation. The channels are simulated as a $4-$mode MMF with SNR $= 20dB$, SNR^e $= 10dB$, and MDL $= 20dB$.

In each plot, the vertical line on the right is the maximum reliable transmission rate between Alice and Bob and the vertical line on the left is the maximum perfect secrecy transmission rate that can be obtained with separate source-channel coding. The horizontal line is the maximum distortion which is also the rate distortion curve from Theorem 7 with no causal information at Eve. The curve obtained from Theorem 8 shows the tradeoff between the transmission rate between Alice and Bob and the distortion forced on Eve with causal information. We see in Fig. 5(a), $p = 0.3$, that with our source-channel coding analysis, we gain a free region for maximum distortion, as if under perfect secrecy, (from the left vertical line to the kink) because we effectively use the redundancy of the source. In Fig. 5(b) with $p = 0.5$, since there is no redundancy in the source, the distortion curve drops immediately after the maximum perfect secrecy rate. Note that the transmission rates are not considered beyond the right vertical lines because they are above the maximum reliable transmission rates and Bob cannot losslesly reconstruct the source sequences.

VI. Conclusion

In this work, we have examined the rate-distortion-based secrecy performance of an insecure MMF communication system. The sender is assumed to have an i.i.d. source sequence which the intended receiver and the eavesdropper both try to reconstruct. Two source-channel coding models with different information availability at the eavesdropper have been considered. We have shown that, when no causal source information is disclosed to the eavesdropper, under a general

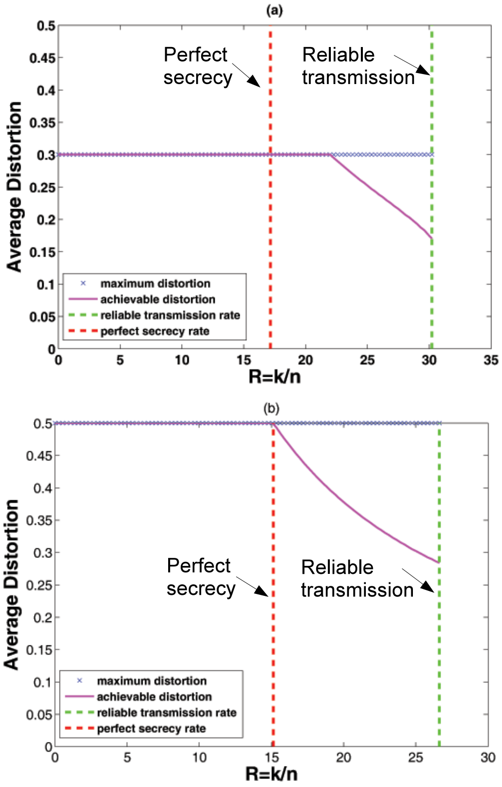


Fig. 5: Achievable distortion-rate curves. On the left is the Bern(0.3) i.i.d. source case and on the right is the Bern(0.5) i.i.d. source case. On the horizontal axes are the symbol/channel use source-channel coding rate and on the vertical axes are the average Hamming distortions.

broadcast channel and any distortion measure, it is possible to send the source at the maximum rate that guarantees lossless reconstruction at the intended receiver while keeping the distortion at the eavesdropper as high as if it only has the source prior distribution. When the past source realization is causally disclosed to the eavesdropper, we have applied the theoretical results in [7] to the particular case of an MMF channel. Numerical results for an i.i.d. Bernoulli source and Hamming distortion have been provided.

Only the theoretical formulation is given to calculate the secrecy outage probability for random MMF under equal full power allocation $Q = I$. The optimality of this power strategy and the statistics for different sets of parameters given in Section IV-B remain open problems. Moreover, in our model, it is required that the intended receiver reconstruct the source losslessly. In a more general setting, one can allow lossy reconstruction of the source at the intended receiver, which is an interesting problem for further research.

Appendix A
Proof of Lemma 2

Given (m_s, m_p), suppose there exists θ_{m_p} such that

$$||P_{Z^n|M_p=m_p,M_s=m_s} - \theta_{m_p}||_{TV} \le \epsilon_n \tag{30}$$

where $\epsilon_n = 2^{-n\beta}$ for some $\beta > 0$. Then we have the following:

$$\begin{aligned}
& ||P_{Z^n|M_p=m_p} - \theta_{m_p}||_{TV} \\
= {} & \sum_{z^n} \left| P_{Z^n|M_p=m_p}(z^n) - \theta_{m_p}(z^n) \right| \qquad (31)\\
= {} & \sum_{z^n} \Bigg| \sum_{m_s} P_{M_s|M_p=m_p}(m_s) P_{Z^n|M_p=m_p,M_s=m_s}(z^n) \\
& - \sum_{m_s} P_{M_s|M_p=m_p}(m_s)\theta_{m_p}(z^n) \Bigg| \\
= {} & \sum_{z^n} \left| \sum_{m_s} \frac{1}{|\mathcal{M}_s|} P_{Z^n|M_p=m_p,M_s=m_s}(z^n) - \sum_{m_s} \frac{1}{|\mathcal{M}_s|}\theta_{m_p}(z^n) \right| \\
\le {} & \sum_{z^n}\sum_{m_s} \frac{1}{|\mathcal{M}_s|} \left| P_{Z^n|M_p=m_p,M_s=m_s}(z^n) - \theta_{m_p}(z^n) \right| \qquad (32)\\
= {} & \sum_{m_s} \frac{1}{|\mathcal{M}_s|} \sum_{z^n} \left| P_{Z^n|M_p=m_p,M_s=m_s}(z^n) - \theta_{m_p}(z^n) \right| \\
\le {} & \sum_{m_s} \frac{1}{|\mathcal{M}_s|} \epsilon_n \qquad (33)\\
= {} & \epsilon_n \qquad (34)
\end{aligned}$$

where (32) follows from triangle inequality and (33) follows from (30). We further have

$$\begin{aligned}
& ||P_{Z^n|M_p=m_p} P_{M_s|M_p=m_p} - P_{Z^n M_s|M_p=m_p}||_{TV} \\
= {} & \sum_{z^n}\sum_{m_s} |P_{Z^n|M_p=m_p}(z^n) P_{M_s|M_p=m_p}(m_s) \\
& - P_{Z^n|M_p=m_p,M_s=m_s}(z^n) P_{M_s|M_p=m_p}(m_s)| \\
= {} & \frac{1}{|\mathcal{M}_s|} \sum_{z^n}\sum_{m_s} |P_{z^n|M_p=m_p}(z^n) \\
& - P_{Z^n|M_p=m_p,M_s=m_s}(z^n)| \\
= {} & \frac{1}{|\mathcal{M}_s|} \sum_{z^n}\sum_{m_s} |P_{Z^n|M_p=m_p}(z^n) - \theta_{m_p}(z^n) \\
& + \theta_{m_p}(z^n) - P_{Z^n|M_p=m_p,M_s=m_s}(z^n)| \\
\le {} & \frac{1}{|\mathcal{M}_s|} \sum_{z^n}\sum_{m_s} (|P_{Z^n|M_p=m_p}(z^n) - \theta_{m_p}(z^n)| \\
& + |P_{Z^n|M_p=m_p,M_s=m_s}(z^n) - \theta_{m_p}(z^n)|) \\
= {} & \frac{1}{|\mathcal{M}_s|} \sum_{m_s} (\sum_{z^n} |P_{Z^n|M_p=m_p}(z^n) - \theta_{m_p}(z^n)| \\
& + \sum_{z^n} |P_{Z^n|M_p=m_p,M_s=m_s}(z^n) - \theta_{m_p}(z^n)|) \\
\le {} & \frac{1}{|\mathcal{M}_s|} \sum_{m_s} (\epsilon_n + \epsilon_n) \\
= {} & 2\epsilon_n
\end{aligned}$$

By applying Lemma 1, we have

$$\begin{aligned}
I(M_s; Z^n|M_p) &= \sum_{m_p} P_{M_p}(m_p) I(M_s; Z^n|M_p = m_p) \\
&\le \sum_{m_p} P_{M_p}(m_p)(-2\epsilon_n \log \frac{2\epsilon_n}{|\mathcal{M}_s|}) \\
&\le 2 \cdot 2^{-n\beta}(nR_s) \qquad (35)
\end{aligned}$$

where (35) goes to 0 as $n \to \infty$.

APPENDIX B
PROOF OF (14)

For each i, we have

$$\begin{aligned} I(S_i; Z^n|M_p) &\leq I(M_s S_i; Z^n|M_p) \\ &= I(M_s; Z^n|M_p) + I(S_i; Z^n|M_s M_p) \\ &\leq \epsilon \end{aligned} \tag{36}$$

for large enough n. (36) follows from strong secrecy of the channel and Fano's inequality. Note that weak secrecy is not sufficient to give us the desired result in our proof. We now define

$$P_i \triangleq P_{S_i Z^n M_p}$$

$$\bar{P}_i \triangleq P_{M_p} P_{S_i|M_p} P_{Z^n|M_p}$$

i.e. $\bar{P}_i$ is the Markov chain $S_i - \square - M_p - \square - Z^n$. By Pinsker's inequality,

$$\begin{aligned} ||P_i - \bar{P}_i||_{TV} &\leq \frac{1}{\sqrt{2}} D(P_i||\bar{P}_i)^{\frac{1}{2}} \\ &= \frac{1}{\sqrt{2}} I(S_i; Z^n|M_p)^{\frac{1}{2}} \\ &\leq \sqrt{\frac{\epsilon}{2}} \end{aligned} \tag{37}$$

$$\begin{aligned} &\min_{t(i,z^n)} \mathbb{E}[d(S_i, t(i, Z^n))] \\ \geq &\min_{t(i,z^n,m_p)} \mathbb{E}[d(S_i, t(i, Z^n, M_p))] \\ \geq &\min_{t(i,z^n,m_p)} \mathbb{E}_{\bar{P}_i}[d(S_i, t(i, Z^n, M_p))] - \delta'(\epsilon) && (38) \\ = &\min_{t(i,m_p)} \mathbb{E}_{\bar{P}_i}[d(S_i, t(i, M_p))] - \delta'(\epsilon) && (39) \\ \geq &\min_{t(i,m_p)} \mathbb{E}[d(S_i, t(i, M_p))] - 2\delta'(\epsilon) && (40) \end{aligned}$$

where (38) and (40) use the fact that P_i and $\bar{P}_i$ are close in total variation from (37); and (39) uses the Markov relation $S_i - \square - M_p - \square - Z^n$ of the distribution $\bar{P}_i$. The technical details can be found in Lemma 2 and 3 from [7]. Averaging over k, we obtain (14).

APPENDIX C
JUSTIFICATION OF THE CONDITION $\max_{(R_s,R_p)\in\mathcal{R}} R_s > 0$

From Theorem 2 or 3, we have

$$\max_{(R_s,R_p)\in\mathcal{R}} R_s > 0$$

is equivalent to

$$I(W; Y|V) - I(W; Z|V) \quad > \quad 0 \tag{41}$$

for some $V - \square - W - \square - X - \square - YZ$. We claim this can be simplified to

$$I(W; Y) - I(W; Z) \quad > \quad 0 \tag{42}$$

for some $W - \square - X - \square - YZ$.

To see (42) $\Rightarrow$ (41), we can simply let $V = \emptyset$. To see (41) $\Rightarrow$ (42), observe that if there exists $V - \square - W - \square - X - \square - YZ$ such that (41) holds, then there has to exist at least one value v such that $I(W; Y|V = v) - I(W; Z|V = v) > 0$. We can redefine the distribution as $P_{\bar{W}\bar{X}\bar{Y}\bar{Z}} \triangleq P_{WXYZ|V=v}$. It can be verified that the Markovity $\bar{W} - \square - \bar{X} - \square - \bar{Y}\bar{Z}$ holds and $P_{\bar{Y}\bar{Z}|\bar{X}} = P_{YZ|X}$.

APPENDIX D
SUFFICIENT CONDITION ON THEOREM 7

From Theorem 1 and Corollary 3, we know that a sufficient condition for the eavesdropper's channel not being less noisy than the intended receiver's channel is

$$\max_{K\in\mathcal{H}^{M\times M}, 0\preceq K\preceq I} \frac{|\text{SNR}K + I|}{|\text{SNR}^e \Psi^e K \Psi^{e\dagger} \Phi + I|} > 1. \tag{43}$$

However, (43) is computationally difficult to verify. If we restrict K to be of the form $K = \Psi^{e\dagger} \Lambda \Psi^e$ where Λ is diagonal with diagonal entries $\lambda_i \in [0, 1]$, then (43) has a much simpler form:

$$\frac{\prod_{i=1}^{M} (1 + \text{SNR}\lambda_i)}{\prod_{i=1}^{M} (1 + \text{SNR}^e \lambda_i \bar{\phi}_i)} > 1. \tag{44}$$

Therefore, if there exists a $j \in \{1, ..., M\}$ such that $\bar{\phi}_j < \frac{\text{SNR}}{\text{SNR}^e}$, we can choose $\lambda_j = 1$ and $\lambda_i = 0$ for $i \neq j$ to satisfy (44).

ACKNOWLEDGMENT

The authors would like to thank Dr. Peter Winzer from Bell Labs, Alcatel Lucent, Dr. Matthieu Bloch and Dr. Rafael Schaefer for fruitful discussions and support of this work.

REFERENCES

[1] P. J. Winzer and G. J. Foschini, "MIMO capacities and outage probabilities in spatially multiplexed optical transport systems," *Opt. Express*, vol. 19, pp. 16680–16696, Aug. 2011.

[2] K. Guan, P. J. Winzer, and E. Soljanin, "Information-theoretic security in space-division multiplexed fiber optic networks," in *Proc. 2012 European Conference and Exhibition on Optical Communication*, p. Tu.3.C.4.

[3] H. D. Ly, T. Liu, and Y. Liang, "Multiple-input multiple-output Gaussian broadcast channels with common and confidential messages," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5477–5487, 2010.

[4] C. Schieler and P. Cuff, "Rate-distortion theory for secrecy systems," *CoRR*, vol. abs/1305.3905, 2013.

[5] H. Yamamoto, "Rate-distortion theory for the Shannon cipher system," *IEEE Trans. Inf. Theory*, vol. 43, no. 3, pp. 827–835, 1997.

[6] P. Cuff, "A framework for partial secrecy," in *Proc. 2010 IEEE Global Telecommunications Conference*, pp. 1–5.

[7] C. Schieler, E. C. Song, P. Cuff, and H. V. Poor, "Source-channel secrecy with causal disclosure," in *Proc. 2012 Allerton Conference on Communication, Control, and Computing*, pp. 968–973.

[8] C. Schieler and P. Cuff, "Secrecy is cheap if the adversary must reconstruct," in *Proc. 2012 IEEE International Symposium on Information Theory*, pp. 66–70.

[9] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 339–348, 1978.

[10] R. F. Wyrembelski and H. Boche, "Strong secrecy in compound broadcast channels with confidential messages," in *Proc. 2012 IEEE International Symposium on Information Theory*, pp. 76–80.

[11] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, pp. 1355–1387, 1975.

[12] U. Maurer and S. Wolf, "Information-theoretic key agreement: from weak to strong secrecy for free," in *Advances in Cryptology EUROCRYPT 2000* (B. Preneel, Ed.), vol. 1807 of *Lecture Notes in Computer Science*, pp. 351–368. Springer Berlin Heidelberg, 2000.

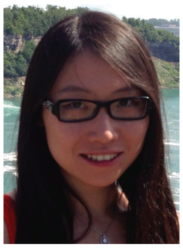
[13] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.

[14] K. Guan, E. C. Song, E. Soljanin, P. J. Winzer, and A. M. Tulino, "Physical layer security in space-division multiplexed fiber optic communications," in *Proc. 2012 IEEE Asilomar Conference on Signals, Systems and Computers*, pp. 654–658.

[15] E. Teletar, "Capacity of multi-antenna Gaussian channels," AT&T Bell Labs, Tech. Rep., 1995.

[16] F. Oggier and B. Hassibi, "The secrecy capacity of the MIMO wiretap channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4961–4972, 2011.

[17] R. F. Schaefer and S. Loyka, "The secrecy capacity of a compound MIMO Gaussian channel," in *Proc. 2013 IEEE Information Theory Workshop.*



Eva C. Song received her B.S. degree in Electrical and Computer Engineering from Carnegie Mellon University, Pittsburgh, PA, in 2010 and her M.A. degree from Electrical Engineering from Princeton University, NJ, in 2012. Since 2010, she has been a Ph.D. candidate in Electrical Engineering at Princeton University. During 2012, she interned at Bell Labs, Alcatel-Lucent, Murray Hill, NJ, to study secrecy in optics communications. Her main research areas include information theoretic secrecy and source coding.

Emina Soljanin received the PhD and MS degrees from Texas A&M University, College Station, in 1989 and 1994, and the European Diploma degree from University of Sarajevo, Bosnia, in 1986, all in Electrical Engineering.

From 1986 to 1988, she worked in the Energoinvest Company, Bosnia, developing optimization algorithms and software for power system control. Aftergraduating from Texas A&M in 1994, she joined Bell Laboratories, Murray Hill, NJ, where she is now a Distinguished Member of Technical Staff in the Mathematics of Networks and Communications research department. Her research interests are in the broad area of communications, information and coding theory, and their applications. In the course of her fifteen year employment with Bell Labs, she has participated in a very wide range of research and business projects. These projects include designing the first distance enhancing codes to be implemented in commercial magnetic storage devices, first forward error correction for Lucent's optical transmission devices, color space quantization and color image processing, quantum computation, link error prediction methods for the third generation wireless network standards, and anomaly and intrusion detection. Her most recent activities are in the area of network and rateless coding. For research in these areas, she has won NSF, DARPA, NAE, and ONR funding, for salary, students, travel, and workshops.

Dr. Soljanin served as a Technical Proof-Reader, 1990-1992, and as the Associate Editor for Coding Techniques, 1997-2000, for the IEEE TRANSACTIONS ON INFORMATION THEORY. She has been serving as a Co-Chair for the DIMACS Special Focus on Computational Information Theory and Coding. She spent 2008 as a visiting researcher at Ecole Polytechnique Federale de Lausanne (EPFL), in Switzerland. Dr. Soljanin is a member of the editorial board of the *Springer Journal on Applicable Algebra in Engineering, Communication and Computing* (AAECP), and a member of the Board of Governors of the IEEE Information Theory Society.

Paul Cuff (S'08–M'10) received the B.S. degree in electrical engineering from Brigham Young University, Provo, UT, in 2004 and the M.S and Ph.D. degrees in electrical engineering from Stanford University in 2006 and 2009. Since 2009 he has been an Assistant Professor of Electrical Engineering at Princeton University. As a graduate student, Dr. Cuff was awarded the ISIT 2008 Student Paper Award for his work titled "Communication requirements for generating correlated random variables," and was a recipient of the National Defense Science and Engineering Graduate Fellowship and the Numerical Technologies Fellowship.

H. Vincent Poor (S'72–M'77–SM'82–F'87) received the Ph.D. degree in EECS from Princeton University in 1977. From 1977 until 1990, he was on the faculty of the University of Illinois at Urbana-Champaign. Since 1990 he has been on the faculty at Princeton, where he is the Michael Henry Strater University Professor of Electrical Engineering and Dean of the School of Engineering and Applied Science. Dr. Poor's research interests are in the areas of information theory, statistical signal processing and stochastic analysis, and their applications in wireless networks and related fields including social networks and smart grid. Among his publications in these areas are the recent books *Smart Grid Communications and Networking* (Cambridge University Press, 2012) and *Principles of Cognitive Radio* (Cambridge University Press, 2013).

Dr. Poor is a member of the National Academy of Engineering, the National Academy of Sciences, and Academia Europaea, and is a fellow of the American Academy of Arts and Sciences, the Royal Academy of Engineering (UK), and the Royal Society of Edinburgh. He received the Marconi and Armstrong Awards of the IEEE Communications Society in 2007 and 2009, respectively. Recent recognition of his work includes the 2010 IET Ambrose Fleming Medal for Achievement in Communications, the 2011 IEEE Eric E. Sumner Award, and honorary doctorates from Aalborg University, the Hong Kong University of Science and Technology, and the University of Edinburgh.

Kyle Guan received the Ph.D. degree in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 2007. From 2007 to 2010, he was a Senior Research Engineer at BAE Systems, Wayne, NJ. He joined Bell Labs, Alcatel-Lucent, NJ, in 2010. His current research interests include network security, energy-efficient network architectures, optical transmission systems, and network optimization.