# Information Theoretic Security and Privacy of Information Systems

*Edited by*
Holger Boche, Ashish Khisti, H. Vincent Poor, and Rafael F. Schaefer

# 3 Secure source coding

Paul Cuff and Curt Schieler

## Abstract

This chapter assumes that a limited amount of secret key and reliable communication are available for use in encoding and transmitting a stochastic signal. The chapter starts at the beginning, with a more general proof of Shannon's "key must be as large as the message" result that holds even for stochastic encoders.

Three directions are explored. First, for lossless compression and perfect secrecy, variable length codes or key regeneration allow for a trade-off between efficiency of compression and efficiency of secret key usage. Second, the relaxation to imperfect secrecy is studied. This is accomplished by measuring the level of secrecy either by applying a distortion metric to the eavesdropper's best possible reconstruction or by considering the eavesdropper's ability to guess the source realization. Finally, an additional relaxation is made to allow the compression of the source to by lossy.

The chapter concludes by showing how the oft-used equivocation metric for information-theoretic secrecy is a particular special case of the rate-distortion theory contained herein.

## 3.1 Introduction

Source coding is the process of encoding information signals for transmission through digital channels. Since efficiency is a primary concern in this process, the phrase "source coding" is often used interchangeably with "data compression." The relationship between source coding and channel coding is that channel coding produces digital resources from natural resources (e.g., a physical medium), and source coding consumes digital resources to accomplish a task involving information, often simply moving it from one point to another.

The channel coding side of information theoretic security is referred to as physical-layer security. This usually involves designing a communication system for a physical wiretap channel, introduced by Wyner in [1], which produces a provably secure digital communication link. Another important challenge in physical-layer security is the production of a secret key based on common observations, such as channel fading parameters or prepared quantum states. Although

a key agreement protocol does not necessarily involve a channel, it is consistent with the spirit of channel coding in that the objective is the production of digital resources. The digital resources generated by physical-layer security or key agreement can be any of the following:

- Common random bits
- Common random bits unknown to an eavesdropper (secret key)
- A reliable digital communication link
- A reliable and secure digital communication link

Secure source coding involves using these digital resources in an efficient way to produce a desired effect with regards to an information signal. The general setting of interest in this chapter is the Shannon cipher system [2] shown in Figure 3.1, where $\mathbf{X}$ represents the information signal, $K$ is a secret key, $M$ is the encoded signal, which is transmitted over a reliable but non-secure digital channel, $\widehat{\mathbf{X}}$ is the reconstructed information, and A and B are the encoder and decoder that we wish to design in order to keep an eavesdropper, Eve, from learning about $\mathbf{X}$ and $\widehat{\mathbf{X}}$. A variation of this setting could incorporate reliable *and secure* communication channels, but the setting of Figure 3.1 provides the necessary elements for fundamental understanding. For simplicity, we will mostly focus on asymptotic results in the case where $\mathbf{X}$ is a sequence of i.i.d. random variables with known distribution. This information signal is referred to herein as the "source."

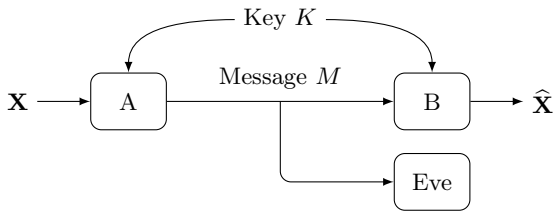


**Figure 3.1** *Shannon cipher system.* An eavesdropper Eve intercepts the communication between the sender and receiver. The sender and receiver share a secret key, which can be used to encrypt the communication.

Notice that secure source coding operates at a layer above the physical layer. The method with which the resources were produced is not the concern of source coding. If, for example, one has confidence in a key that was produced and secured by some cryptographic technique, a feasible implementation would be to encode the source using that key rather than an information-theoretic secret key. However, to derive tight results, we assume that secure channels and secret keys are information-theoretically secure.

One interesting feature of secure communication is the role of stochasticity in the encoder or decoder. The intentional inclusion of randomness in the encoding

or decoding process can lead to superior designs in spite of the reliability issues it introduces. This effect is discussed throughout the chapter.

**Box 3.1** Stochastic encoder and stochastic decoder

A stochastic encoder or stochastic decoder uses internal (i.e. local) randomness while performing its function. The encoder may not always produce the same message even when encountering the same source value and the same key. Likewise, the decoder may not produce the same reconstruction even if the message and the key are the same.

This chapter begins with the setting that demands the most in terms of digital resources: the source must be recovered without loss and the secrecy must be perfect. This setting is addressed in Section 3.2. Then, just as classic source coding must be generalized to answer questions such as how to encode continuous-valued sources, which cannot be encoded without loss, so a comprehensive view of secure source coding must address how to best secure a source even when the resources needed for perfect secrecy are either not available or too expensive. Each of these generalizations provides rich new theory, highlighted throughout the chapter. In Section 3.3, several notions of imperfect secrecy are introduced, including causal disclosure, repeated and single guessing, and henchmen. Section 3.4 considers lossy compression and imperfect secrecy.

## 3.2 Lossless compression, perfect secrecy

When encoded messages are transmitted over a non-secure digital channel, secrecy can be achieved by use of a secret key (represented as $K$ in Figure 3.1), known to the transmitter and receiver, but unknown to an eavesdropper.

The most ideal secrecy, "perfect secrecy," requires the transmitted message $M$ to be independent of the source $\mathbf{X}$. That is, the conditional distribution of the message given the source, $P_{M|\mathbf{X}}$, does not actually depend on the source $\mathbf{X}$. This simple definition is more general that other notions of secrecy presented in this chapter because it is well-defined even if the source is not modeled as random.

Even before the advent of information theory, it was already known that perfect secrecy can be achieved by using the Vernam cipher with a one-time-pad.

**Box 3.2** One-time pad

Message bits and secret key bits are aligned, and the transmission is computed by bit-wise exclusive-or operations. The key should be uniformly distributed and used only once.

The one-time-pad requires that the space of secret key values $\mathcal{K}$, over which

the key is distributed uniformly at random, be at least as large as the space of possible source values $\mathcal{X}$. Shannon showed this to be necessary in general.

THEOREM 3.1 ([2, Thm. 6]) *Perfect secrecy requires* $|\mathcal{K}| \geq |\mathcal{X}|$.

*Proof* Achievability is accomplished with the one-time-pad.

Shannon's proof of the converse in [2] assumed that the encoder is deterministic. However, we will see that in general it can be beneficial to allow the encoder to be stochastic. For example, this allows for more efficient key usage in variable-length-coding and key regeneration, as will be discussed in this section.[1]

A feature of perfect secrecy is that it is not affected by any assumption about the distribution of the source. For simplicity, let $\mathbf{X}$ be random with a support that covers the whole space $\mathcal{X}$. Let $\mathcal{M}_{x,k}$ be the conditional support of the message given the source and key values, and similarly for $\mathcal{M}_k$ and $\mathcal{M}_x$.

$$
\begin{aligned}
|\mathcal{M}| &\geq \max_k |\mathcal{M}_k| && (3.1)\\
&= \max_k \left| \bigcup_x \mathcal{M}_{x,k} \right| && (3.2)\\
&\overset{(a)}{=} \max_k \sum_x |\mathcal{M}_{x,k}| && (3.3)\\
&\geq \frac{1}{|\mathcal{K}|} \sum_{x,k} |\mathcal{M}_{x,k}| && (3.4)\\
&= \frac{|\mathcal{X}|}{|\mathcal{K}|} \frac{1}{|\mathcal{X}|} \sum_{x,k} |\mathcal{M}_{x,k}| && (3.5)\\
&\geq \frac{|\mathcal{X}|}{|\mathcal{K}|} \min_x \sum_k |\mathcal{M}_{x,k}| && (3.6)\\
&\geq \frac{|\mathcal{X}|}{|\mathcal{K}|} \min_x \left| \bigcup_k \mathcal{M}_{x,k} \right| && (3.7)\\
&= \frac{|\mathcal{X}|}{|\mathcal{K}|} \min_x |\mathcal{M}_x| && (3.8)\\
&\overset{(b)}{=} \frac{|\mathcal{X}|}{|\mathcal{K}|} |\mathcal{M}|, && (3.9)
\end{aligned}
$$

where (a) is a consequence of decodability, which requires that $\{\mathcal{M}_{x,k}\}_x$ are mutually exclusive for any $k$, and (b) is due to perfect secrecy, which requires that the conditional distribution of the message $M$ be constant for all source values. □

*Remark 3.1* If the encoder is deterministic, then $|\mathcal{M}_{x,k}| = 1$ for all $m$ and $k$. Therefore, (3.4) shows that in fact $|\mathcal{K}| \geq |\mathcal{M}| \geq |\mathcal{X}|$ is necessary in that case.

[1] In Wyner's wiretap channel [1], which is outside of the scope of this chapter, a stochastic encoder is necessary to achieve the secrecy capacity.

*Remark 3.2* The bit-wise exclusive-or implementation of the one-time-pad described in Box 3.2 is common, but in general, for a key space of size $|\mathcal{K}|$, a single modulo $|\mathcal{K}|$ addition of the message and the key, after mapping them both to integer values, is an equally effective implementation of perfect secrecy that works even if the size of the key space is not a power of two.

### 3.2.1 Discrete memoryless source

When the source is modeled as a stochastic process, there is opportunity for compression. Assume the source is an i.i.d. sequence distributed according to $P_X$. That is:

$$\mathbf{X} = X^n \triangleq (X_1, \ldots, X_n). \tag{3.10}$$

In the block encoding framework, the definition of lossless encoding can be relaxed to near-lossless by requiring the probability that the reconstruction does not equal the source sequence to be arbitrarily small. In this setting, the message size and secret key size are parametrized as rates. That is,

$$|\mathcal{M}| = 2^{nR}, \tag{3.11}$$
$$|\mathcal{K}| = 2^{nR_K}. \tag{3.12}$$

Under this relaxed notion of lossless compression, the minimum encoding rate $R$ needed is the entropy of the source distribution, which can be achieved by enumerating the set of typical sequences.

Compression and secrecy can be combined by first compressing the source and then applying a one-time-pad to the compressed message[2]. Indeed, this approach is optimal in the lossless setting, and there is no conflict between the two resources, communication rate and key rate.

THEOREM 3.2 *The closure of the set of rate pairs $(R, R_K)$ for which near-lossless compression and perfect secrecy can be simultaneously achieved is all pairs satisfying*

$$R \geq H(X), \tag{3.13}$$
$$R_K \geq H(X). \tag{3.14}$$

*Proof* Achievability is accomplished by first compressing and then applying the one-time-pad.

The converse for (3.13), well-known from lossless compression, is not affected by a secret key.

[2] This is also mentioned in [2], and the proof is omitted.

The converse for (3.14) is as follows:

$$\begin{aligned} nR_K &\geq H(K) && (3.15)\\ &\geq I(K; X^n|M) && (3.16)\\ &= I(K, M; X^n) - I(M; X^n) && (3.17)\\ &\geq I(\widehat{X}^n; X^n) - I(M; X^n) && (3.18)\\ &= H(X^n) - H(X^n|\widehat{X}^n) - I(M; X^n) && (3.19)\\ &= nH(X) - H(X^n|\widehat{X}^n) - I(M; X^n). && (3.20) \end{aligned}$$

The proof is completed by dividing both sides by $n$ and noting that Fano's inequality makes $\frac{1}{n}H(X^n|\widehat{X}^n)$ vanish as the error probability vanishes. The other term in (3.20), $I(M; X^n)$, is zero due to the perfect secrecy requirement. □

*Remark 3.3* Just as lossless compression is relaxed to near-lossless compression, perfect secrecy can also be relaxed to near-perfect secrecy. However, this change does not affect the fundamental limits of Theorem 3.2. Notice that the last term in (3.20) vanishes as long as $I(M; X^n) \in o(n)$, a condition referred to as "weak secrecy" in the literature.

*Remark 3.4* The proof of Theorem 3.2 actually gives a stronger statement than Theorem 3.1. If we assume exact lossless compression and exact perfect secrecy, then (3.15) and (3.20) yield $H(K) \geq nH(X)$. Furthermore, exact lossless compression and exact perfect secrecy are not affected by the source distribution, so the bound is true for any distribution on the source space. By choosing the uniform distribution, we obtain $H(K) \geq n \log |\mathcal{X}| = \log |\mathcal{X}^n|$.

### Variable length coding

We now return to exact lossless compression but take advantage of the source distribution in a different way. Consider an encoder that varies the length of communication or secret key usage depending on the source value. This differs from the fixed-length block encoding model described by (3.11) and (3.12). In a variable-length-coding setting, the resource usage is captured by the expected value of the length.

Consider the encoding of individual source symbols $X \sim P_X$. Reliable encoding requires two parts. The decoder should be able to uniquely decode without punctuation between the encoded symbols, and there should be a causal stopping rule by which the decoder is able to identify the number of key bits needed for decoding. In secrecy settings, pessimism is often prudent, so let us demand perfect secrecy even if an eavesdropper is given punctuation between the symbols. Such would be the case if there were detectable delays between symbol transmissions or if just a single symbol were transmitted.

It is immediately apparent that in order to achieve perfect secrecy, not even the codeword length can reveal information about the source. Thus, the codeword length, if it varies, must be a random variable independent of the source. Un-

surprisingly, then, only codes with fixed-length codewords ultimately need to be considered. However, the variable length of the secret key can still be beneficial.

Prefix codes, of which the Huffman code has the shortest expected length, play an important role in lossless compression by providing unique decodability. It is shown in [3] and [4] that prefix codes also play an important role in this secrecy context. However, there is a tension between communication rate and key rate, which is expressed in the following achievability theorem.

THEOREM 3.3 ([3, Theorem 1],[4]) *A set of achievable rate pairs $(R, R_K)$ for which lossless compression and perfect secrecy can be simultaneously achieved using variable length codes is the convex hull of the following set:*

$$\bigcup_{\mathcal{C}\text{: prefix code}} \left\{ (R, R_K) : \begin{array}{c} R \geq l_{\mathcal{C},\max}, \\ R_K \geq \mathbb{E} l_{\mathcal{C}}(X). \end{array} \right\}. \tag{3.21}$$

*Proof* To achieve these rates, the encoder uses a prefix code and applies a one-time pad using the secret key. Then the encoder appends uniformly distributed random bits to make the transmission equal in length to the longest codeword. The decoder recovers one bit at a time using the secret key and the received transmission. Because of the prefix condition, the decoder can identify when the codeword is complete and declare a stop to the secret key usage. From the point of view of an eavesdropper, every transmission of length $l_{\max}$ is equally likely no matter which source symbol is encoded.

The convex hull is achieved by time-sharing between codes. □

*Remark 3.5* There is tension between the two rates, with the key rate always the smaller of the two. The best code for the key rate is the Huffman code, which achieves a key rate within one bit of the entropy of the source, but which may unfortunately have a long maximum codeword length. At the other extreme of the trade-off, the fixed length code gives $R = R_K = \lceil \log |\mathcal{X}| \rceil$, which corresponds to Theorem 3.1.

*Remark 3.6* The savings in $R_K$ stated in Theorem 3.3 are achieved by using a stochastic encoder. If the encoder is required to be deterministic, then instead one might use secret key to pad each codeword to make them the same length. This is inefficient, and in fact the optimal rates reduce back to those implied by Theorem 3.1.

### Key regeneration

Even in the block encoding model with fixed length communication and a fixed-length key, there is some sense in which less key rate is needed than Theorem 3.1 implies. At the expense of a higher communication rate, and by using a stochastic encoder, it may be that for some realizations of the source not all of the secrecy of the key is exhausted, in the sense that a fresh key can be extracted by the encoder and decoder for future use without compromising security. This gives a way to

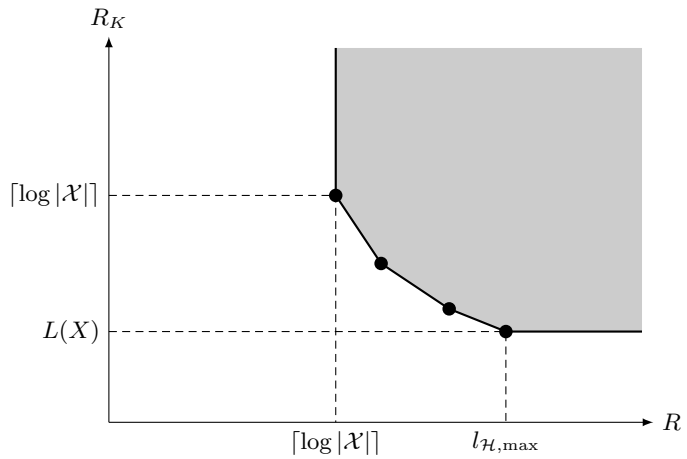


**Figure 3.2** Region of achievable rate pairs for variable-length coding given by Theorem 3.3. Here, $L(X)$ denotes the average codeword length of a Huffman code (i.e, $\min_{\mathcal{C}} \mathbb{E} l_{\mathcal{C}}(X)$), and $l_{\mathcal{H},\max}$ is the maximum codeword length of the Huffman code.

take advantage of the known source distribution without compromising exact lossless compression or exact perfect secrecy, even in the fixed-length setting.

The easiest way to see this phenomenon is to apply the variable-length encoding scheme used for Theorem 3.3, one symbol at a time, to the fixed-length block encoding setting. This is done with a slight modification. The communication length is already constant in that scheme, but the key usage is variable and must be converted to a fixed-length scheme. To do this, let the key length be as long as the communication (i.e. the maximum codeword length of the prefix code), and perform a one-time-pad on the entire transmission, which includes both the codeword from the prefix code and the random padding that follows it. Now this is a fixed-length scheme with perfect secrecy and lossless compression—both the communication and the secret key are as long as the longest codeword. However, notice that the padding itself can be considered a new random secret key. Both parties can compute it, and it is independent of all previous source symbols and transmissions.

In [4] a so-called partition code is used to achieve the same effect, though the extraction of the new secret key is not as straightforward. The overall conclusion is analogous to Theorem 3.3 though. If the communication rate is to be minimized, then the key rate and the communication rate should be the same and equal to $\log |\mathcal{X}|$, corresponding to Theorem 3.1. There is no need for a stochastic encoder in that case. On the other hand, with a stochastic encoder and a higher transmission rate it is possible to regenerate a new secret key so that the usage of secret key rate is reduced to the entropy of the source. This results in essentially the same picture as Figure 3.2 but without the round-off inefficiencies that

arise from symbol-by-symbol encoding. That is, the extreme rates are $\log|\mathcal{X}|$ and $H(X)$ rather than $\lceil \log|\mathcal{X}| \rceil$ and the expected Huffman codeword length.

## 3.3 Lossless compression, imperfect secrecy

It is natural to investigate the quality of source coding even when it falls between the extremes of perfect recovery and no information. In a secure source coding system, a high quality signal should be received by the intended recipient, and only a low quality signal should be obtainable by an eavesdropper.

### 3.3.1 Rate-distortion theory

It has become standard practice in information theory, beginning with Shannon's original paper [5], to measure the quality of an encoding by the average distortion that a reconstruction incurs. That is, for any distortion function $d(x, \hat{x})$, where $x$ represents the value of the source and $\hat{x}$ represents the value of the reconstruction, consider the average distortion over a block of length $n$ to be

$$d(x^n, \hat{x}^n) = \frac{1}{n} \sum_{t=1}^{n} d(x_t, \hat{x}_t). \tag{3.22}$$

The study of the trade-off between this average distortion and the rate of compression is the celebrated *rate-distortion theory*.

In a secrecy system, we can explore the same trade-off but with respect to the eavesdropper. To begin with, let us consider the case where the encoding is lossless for the intended receiver (more accurately, it is near-lossless). Lossy compression will be explored in Section 3.4. Unlike the previous section, we will no longer require perfect secrecy. Instead, secrecy will be measured by the average distortion that an eavesdropper incurs if he tries to reconstruct the source.

One key feature in the theory of secure source coding is the role that past information about the values of the source can play in helping an eavesdropper reconstruct a signal. We refer to this as *causal disclosure*.

**Box 3.3** Causal disclosure

The eavesdropper obtains the past values of the source, or noisy observations of them, before making its reconstruction of the source at the present time. These observations are modeled as outputs $W$ of a memoryless channel $P_{W|X}$.

Causal disclosure means that at time index $t$, as the eavesdropper attempts to reconstruct the source symbol $X_t$, he has access to the eavesdropped message $M$ and noisy observations $W^{t-1} = (W_1, \ldots, W_{t-1})$ of all of the past source symbols $X^{t-1} = (X_1, \ldots, X_{t-1})$, obtained as side information. The distribution of the

**Box 3.4** Full causal disclosure

Full causal disclosure is the worst-case assumption that $W = X$. Secrecy with respect to full causal disclosure is robust—that is, the analysis will be valid no matter what causal side information is obtained by an eavesdropper.

noise $P_{W|X}$ must be modeled or assumed. In the full causal disclosure setting, there is no noise, so the eavesdropper may use $M$ and the past source symbols $X^{t-1}$ themselves while forming his estimate $\widehat{X}_t$.

Causal disclosure does not play a role in rate-distortion theory outside of the context of secrecy—if causal disclosure is available to the legitimate receiver, it does not change the rate-distortion theorem. But it fundamentally changes the nature of secrecy. If one thinks of the sequence of source symbols as a time sequence, then full causal disclosure, where $W = X$, is akin to the known-plaintext setting in cryptography. Without causal disclosure (e.g., if the channel is $P_{W|X} = P_W$), the encoder can essentially reuse the secret key indefinitely to force a high level of distortion upon the eavesdropper with a negligible key rate. This gives the impression of perfect secrecy for free, but it is really just an artifact of a presumptuous model that assumes an eavesdropper does not obtain additional side information.

DEFINITION 3.4 Given a memoryless source $\{X_t \sim P_X\}$, a causal disclosure channel $P_{W|X}$, and a distortion function $d(x, z)$, a rate-distortion triple $(R, R_K, D)$ is achievable if for all $\epsilon > 0$ there exist a blocklength $n$ and an encoder and decoder at compression rate $R$ and key rate $R_K$ as described in (3.11) and (3.12) such that

$$\mathbb{P}(X^n \neq \widehat{X}^n) < \epsilon, \tag{3.23}$$

$$\min_{\{Z_t = z_t(M, W^{t-1})\}_t} \mathbb{E}\, d(X^n, Z^n) \geq D. \tag{3.24}$$

THEOREM 3.5 ([6, Corollary 1]) *The closure of achievable rate-distortion triples $(R, R_K, D)$ for lossless compression with security robust to causal disclosure is the set of triples satisfying*

$$\bigcup_{U-X-W} \left\{ (R, R_K, D) : \begin{array}{r} R \geq H(X) \\ R_K \geq I(W; X|U) \\ D \leq \min_{z(u)} \mathbb{E}\, d(X, z(U)) \end{array} \right\}. \tag{3.25}$$

*Achievability proof sketch* To achieve the above rates, two messages are transmitted. For the first message, a codebook is generated from the $P_U$ distribution to form a covering of the source space. The rate of this codebook is $I(X;U)$. Note that the key is not used in the encoding of the first message. The second message, at rate $H(X|U)$, is simply a uniformly distributed random mapping

from each source sequence and key pair $(x^n, k)$. These two messages, together with the key, allow for decoding of the source.

The case of full causal disclosure admits a simple secrecy argument. In that case, the bound on $R_K$ in the theorem simplifies to $R_K \geq H(X|U)$, which is the rate of the second message. Therefore, a one-time-pad can be used on the second message, which is effectively what the previously described encoding will yield anyway. Consequently, the eavesdropper receives exactly one piece of information, which is the codeword given by the first message.

The analysis is greatly assisted by the use of a likelihood encoder. Let us refer to this codeword associated with the first message as $\bar{u}^n$, to which the eavesdropper has access. Due to the likelihood encoder, the posterior distribution of the source sequence given the first message is approximately

$$P_{X^n|M_1}(x^n|m_1) \approx \prod_{t=1}^{n} P_{X|U}(x_t|\bar{u}_t). \tag{3.26}$$

At least in the case of full causal disclosure, it is easy to see that this will lead to the distortion stated in the theorem, since the only meaningful information that the eavesdropper obtains is $\bar{u}^n$. The second message is fully protected by the key, and information leaked through causal disclosure has been sterilized by this encoding, as indicated in (3.26), rendering it useless to the eavesdropper. □

*Remark 3.7* The variable $U$ can be interpreted as a part of the encoding that is not secured due to lack of secret key resource. This leaked information is designed to be as useless as possible under the constraints of Theorem 3.5. In the case of full causal disclosure, the rest of the encoding is fully secured.

This decomposition of the encoding into a secured part and a part that is not secured is fortuitous. The encoder has two digital outputs. For the one that must be secured, this can be accomplished by using a one-time-pad or even by using a secure digital channel. This opens the door to implement this source coding with cryptographic resources.

*Remark 3.8* If there is no causal disclosure, then the theorem puts no lower bound on the key rate, and the optimal variable choice is $U = \emptyset$. As discussed previously, optimal distortion is forced on the eavesdropper without needing a positive secret key rate.

*Remark 3.9* Theorem 3.5 does not change even if the disclosure is assumed to be non-causal, meaning that the disclosure is available at all times except the one that is to be reconstructed. This is relevant to source coding settings where the sequence is not a time sequence and causality is unimportant.

*Remark 3.10* Theorem 3.5 does not change if (3.24) is changed to

$$\max_{\{Z_t = z_t(M, W^{t-1})\}_t} \mathbb{P}\left(d(X^n, Z^n) < D\right) < \epsilon. \tag{3.27}$$

*Remark 3.11* The optimization involved in finding the boundary points in Theorem 3.5 simplifies to a linear program.

**Example 3.1** *Hamming distortion.* Let distortion at the eavesdropper be measured by Hamming distortion (i.e. probability of symbol error), and consider full causal disclosure (i.e., set $W = X$). For any source distribution $P_X$, apply Theorem 3.5 to find the maximum distortion that can be forced on the eavesdropper as a function of the secret key rate.

Solution

Define the function $\phi(\cdot)$ as the linear interpolation of the points $(\log k, \frac{k-1}{k}), k \in \mathbb{N}$. This function $\phi(\cdot)$, which is the upper boundary of the convex hull of the set of all possible entropy–error-probability pairs for arbitrary distributions[3], is found in [6] to be the main quantity of interest for this setting. Also, define

$$d_{\max} = 1 - \max_x P_X(x). \tag{3.28}$$

The maximum distortion as a function of the key rate, shown in Figure 3.3, is

$$D(R_K) = \min\{\phi(R_K), d_{\max}\}. \tag{3.29}$$

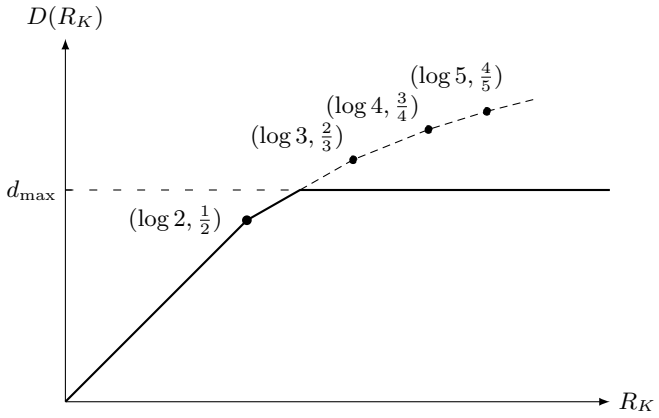


**Figure 3.3** Illustration of $D(R_K)$.

The corner-points of $\phi(\cdot)$ are achieved by constructing $P_{X|U}$ to be a uniform distribution over $k$ values of $X$ for each values of $U$. This can be shown to be possible as long as $\frac{k-1}{k} \leq d_{\max}$.

The two quantities $\phi(R_K)$ and $d_{\max}$ are complementary upper bounds on the distortion. The source distribution $P_X$ only plays a role through $d_{\max}$, which is the best distortion one could hope for even with perfect secrecy, and the key rate is represented in $\phi(R_K)$.

[3] Fano's inequality relates to the lower boundary.

### 3.3.2 Henchman

Another approach to characterizing partial secrecy is to consider how much extra information would be needed for an eavesdropper to form a good reconstruction of the source. In the henchman setting [7], this is formalized by supposing that a henchman is assisting the villainous eavesdropper. The henchman is aware of everything, including the source sequence realization, the ciphertext, and the key value. After secure communication is complete, the henchman uses rate-limited digital communication to help the eavesdropper achieve low distortion. This is illustrated in Figure 3.4.

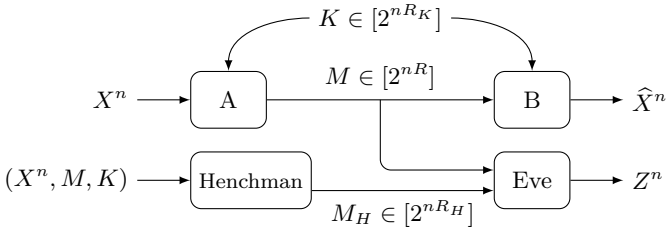


**Figure 3.4** *The henchman problem.* A rate-limited henchman has access to the source sequence and the public message. The eavesdropper produces a reconstruction sequence $Z^n$ based on the public message and the side information from the henchman.

Secure source coding should inflict high distortion on an eavesdropper in spite of side information. In the previous subsection, causal disclosure modeled the side information. In the henchman setting, the side information is the henchman's message, which is rate limited but not causal. It may be an appropriate worst-case analysis when decoding is not needed in real-time or when the source sequence is not a time sequence.

A distortion level $D$ is considered achieved if the eavesdropper's expected distortion is above $D$ no matter which rate $R_H$ communication scheme is devised with the henchman, as defined below.

DEFINITION 3.6 Given a memoryless source $\{X_t \sim P_X\}$ and a distortion function $d(x, z)$, a rate-distortion quadruple $(R, R_K, R_H, D)$ is achievable if for all $\epsilon > 0$ there exist a blocklength $n$ and an encoder and decoder at compression rate $R$ and key rate $R_K$ as described in (3.11) and (3.12) such that

$$\mathbb{P}(X^n \neq \widehat{X}^n) < \epsilon, \tag{3.30}$$

$$\min_{\substack{Z^n = z^n(M, M_H),\\ M_H = m_H(X^n, M, K),\\ |\mathcal{M}_H| = 2^{nR_H}}} \mathbb{E}\, d(X^n, Z^n) \geq D. \tag{3.31}$$

THEOREM 3.7 ([7, Theorem 1]) *The closure of achievable rate-distortion tuples $(R, R_K, R_H, D)$ for lossless compression with security robust to worst-case side*

*information is the set of tuples satisfying*

$$R \geq H(X), \tag{3.32}$$

$$D \leq D(R_H) \cdot \mathbb{1}\{R_K > R_H\}, \tag{3.33}$$

*where* $\mathbb{1}$ *is the indicator function and* $D(r)$ *is the point-to-point distortion-rate function:*

$$D(r) \triangleq \min_{P_{Z|X}: r \geq I(X;Z)} \mathbb{E}\, d(X, Z). \tag{3.34}$$

*Proof sketch* The converse is trivial. Consider that the henchman and the eavesdropper always have the option of compressing the source using a rate-distortion code or revealing the key if $R_H$ is sufficiently high.

Achievability is proven using a random codebook construction. The important step in the secrecy analysis is to show that for large $n$ a random sequence that is uniformly distributed on a random set (generated according to the source distribution) is with high probability not compressible by the henchman beyond the rate-distortion function for memoryless sources, as long as the compression rate $R_H$ is less than the exponential rate of the set size, which is $R_K$ in this context. □

*Remark 3.12* If the henchman's rate is below the secret key rate, then the system effectively attains perfect secrecy. The henchman can only describe the source sequence as efficiently as rate-distortion theory allows. Conversely, if the henchman has enough rate to reveal the key, then the system provides no secrecy. Thus, in this model, secrecy is entirely dependent on whether or not the eavesdropper can obtain as much side information as the rate of the key.

*Remark 3.13* Another interpretation of the henchman setting is that of list decoding. The eavesdropper produces a list of reconstructions of size $2^{nR_H}$, and performance is measured with respect to the best reconstruction in the list.

*Remark 3.14* Theorem 3.7 does not change if (3.31) is changed to

$$\max_{\substack{Z^n = z^n(M, M_H), \\ M_H = m_H(X^n, M, K), \\ |\mathcal{M}_H| = 2^{nR_H}}} \mathbb{P}\left(d(X^n, Z^n) < D\right) < \epsilon. \tag{3.35}$$

### 3.3.3 Repeated guessing

Another model for partial secrecy has the eavesdropper repeatedly guessing until he correctly identifies the source sequence [8]. Secrecy can be quantified by the expected number of guesses required (or any moment thereof). This setting addresses a situation where a password is being guessed and either confirmed or rejected upon each guess.

The dominant terms in the guessing moments do not come from source sequences that are statistically typical. Instead, a low probability set contributes almost all of the moment.

DEFINITION 3.8 Let $G(X^n|M)$ be the number of guesses made by the eavesdropper in order to guess $X^n$ correctly upon observing message $M$, according to the optimal guessing strategy which orders the guesses from most probable to least probable.

Given a memoryless source $\{X_t \sim P_X\}$, a *$\rho$th-moment guessing exponent* $E_\rho$, for $\rho > 0$, is achievable if for arbitrarily large $n$ there exist an encoder and decoder that achieve exact lossless compression and use key rate $R_K$ as described in (3.12) such that

$$\mathbb{E}\left[G(X^n|M)^\rho\right] \geq 2^{nE_\rho}. \tag{3.36}$$

THEOREM 3.9 ([8, Theorem 1]) *The supremum of achievable $\rho$th moment guessing exponents is*

$$E_\rho^* = \max_Q \{\min\{R_K, H(Q)\} \cdot \rho - D(Q\|P_X)\}. \tag{3.37}$$

*Proof* The method of types is appropriate for this analysis. Since each type contributes exponentially to the moment, and there are only a polynomial number of types, it can be argued that the guessing exponent remains unchanged even if the eavesdropper is told the type of $X^n$. Then, to calculate the moment, for each type we multiply the probability of the type by the conditional moment of guesses to find $X^n$ within the type.

The relative entropy term $D(Q\|P_X)$ accounts for the probability of the type $Q$.

A priori, the distribution of source sequences within a type $Q$ is uniform, and the size is $2^{nH(Q)}$ to first order in the exponent. With proper encoding, the posterior distribution of the source sequence given the type and the message is uniformly distributed over a set equal to the size of the key space, unless the key space is larger than the size of the type. One way to accomplish this is to enumerate the sequences and the key values and have the encoding simply add them modulo the size of the type.

Ultimately, the eavesdropper must either guess the uniformly distributed key value or guess the sequence uniformly from the type, whichever is easier. The $\rho$th moment for doing this is asymptotically equivalent to the exponential size of the set to the $\rho$th power. □

*Remark 3.15* Theorem 3.9 assumes that the encoding must be exact lossless compression. If near-lossless encoding is permitted, then perfect secrecy is achieved under this metric even without any key.[4] Non-typical sequences can all be encoded to one message value, and this low probability event will give the same guessing moment as perfect secrecy. The expression for the guessing exponent becomes

$$E_\rho^* = \max_Q \{H(Q) \cdot \rho - D(Q\|P_X)\} = \rho H_{\frac{1}{\rho+1}}(X), \tag{3.38}$$

[4] Similarly, if a variable key rate is used, then a negligible expected key rate is needed to achieve perfect secrecy under this metric, even while achieving exact lossless encoding.

where $H_\alpha(\cdot)$ is the Rényi entropy of order $\alpha$.[5]

*Remark 3.16* If $R_K \le H(X)$, then $E_\rho^* = \rho R_K$ according to Theorem 3.9, and the eavesdropper is essentially guessing the key. The effectiveness of the key begins to diminish at rates higher than this, and the formula in the theorem begins to be interesting.

Recall that when $R_K > H(X)$ exact perfect secrecy is achievable for near-lossless compression, according to Theorem 3.2. On the other hand, $R_K \ge \log|\mathcal{X}|$ is needed for perfect secrecy in the exact lossless compression setting according to Theorem 3.1. This repeated guessing setting is focused on the in-between region. Here exact lossless compression is required, but the notion of perfect secrecy is relaxed.

*Remark 3.17* By this guessing moment metric, secret key is no longer effective once the key rate surpasses $R_K > \bar{H}\left(P_X^{\frac{1}{\rho+1}}\right)$, where the bar over $H$ indicates that the argument is first normalized to be a probability distribution.

*Remark 3.18* A uniformly random code construction will achieve this secrecy performance with high probability.

### 3.3.4 Best guess

If the eavesdropper makes a single guess of the source sequence, we would hope that the probability of correctly guessing is exponentially small. This probability is proposed in [10] as a measure of partial secrecy. In fact, it turns out that under optimal encoding, the eavesdropper will be forced to either guess the value of the key or guess the a priori most likely source sequence, whichever gives the higher probability of success.

DEFINITION 3.10 Given a memoryless source $\{X_t \sim P_X\}$, a *best-guess exponent* $E_G$ is achievable if for arbitrarily large $n$ there exist an encoder and decoder that achieve lossless compression and use key rate $R_K$ as described in (3.12) such that

$$\max_{x^n,m} P_{X^n|M}(x^n|m) \le 2^{-nE_G} \tag{3.39}$$

THEOREM 3.11 ([10]) *The supremum of achievable best-guess exponents is*

$$E_G^* = \min\{R_K, H_\infty(X)\}, \tag{3.40}$$

*where*

$$H_\infty(X) = \log \frac{1}{\max_x P_X(x)} \tag{3.41}$$

*is the min-entropy or Rényi entropy of order $\infty$.*

[5] This matches the guessing exponent derived in [9, Proposition 5] for guessing a source sequence in the absence of a message.

*Proof* The converse is trivial. The eavesdropper may always choose the better of two options: guess the key value without taking into account the source distribution, or guess the most likely source sequence while ignoring the message. These are the two upper bounds that the theorem achieves.

For achievability, the method of types is again convenient. First use a negligible amount of key to keep the type secret, then encode within each type exactly as was done for the proof of Theorem 3.9.

For a sequence of type $Q$, first consider what happens if the key space is smaller than the size of the type. Then the posterior probability of any one of these sequences is equivalent to the probability of correctly guessing both the type and the key value. This upper-bounds the probability by $2^{-nR_K}$. On the other hand, if the key space is larger than the size of the type, then perfect secrecy is achieved for each of the sequences in that type, and

$$P_{X^n|M}(x^n, m) = P_{X^n}(x^n) \leq 2^{-nH_\infty(X)}. \tag{3.42}$$

□

*Remark 3.19* The statement of Theorem 3.11 does not uncover any complexities or surprises in either the solution formula or the encoding needed to achieve it. Indeed, a uniform random code construction will achieve this secrecy performance with high probability. Perhaps the most interesting observation is the similarity of the mathematics between Theorems 3.9 and 3.11.

Notice that $E_\rho^*$ evaluated at $\rho = -1$ is equal to the negative of $E_G^*$. That is,

$$\begin{aligned}
-E_{-1}^* &= -\max_Q\{-\min\{R_K, H(Q)\} - D(Q\|P_X)\} && (3.43)\\
&= \min_Q\{\min\{R_K, H(Q)\} + D(Q\|P_X)\} && (3.44)\\
&= \min\{\min_Q\{R_K + D(Q\|P_X)\},\ \min_Q\{H(Q) + D(Q\|P_X)\}\} && (3.45)\\
&= \min\left\{R_K, \min_Q \mathbb{E}_Q \log\frac{1}{P_X(X)}\right\} && (3.46)\\
&= \min\{R_K, H_\infty(X)\}. && (3.47)
\end{aligned}$$

This is not a coincidence. Even though the repeated guessing problem definition constrains $\rho > 0$, it can be extended to include $\rho < 0$ with a change in the direction of inequality in the definition of achievability:

$$\mathbb{E}[G(X^n|M)^\rho] \leq 2^{nE_\rho}. \tag{3.48}$$

This change is made because $(\cdot)^\rho$ is now monotonically decreasing on the positive domain when $\rho < 0$.

Under this change to Definition 3.8, when we consider $\rho = -1$, we get Definition 3.10. This is verified by the following inspection. Consider any random variable $W$ on the positive integers with a monotonically non-increasing probability

mass function ($W$ will take the role of $G(X^n|M)$), and define $p_{\max} = P_W(1)$.

$$\mathbb{E}[W^{-1}] \geq p_{\max}. \tag{3.49}$$

$$\mathbb{E}\left[W^{-1}\right] \leq p_{\max} \sum_{k=1}^{\left\lceil \frac{1}{p_{\max}} \right\rceil} \frac{1}{k} \tag{3.50}$$

$$\leq p_{\max} \ln\left(\left\lceil \frac{1}{p_{\max}} \right\rceil + 1\right). \tag{3.51}$$

$$\tag{3.52}$$

Therefore, $\mathbb{E}[W^{-1}]$ is exponentially equivalent to $p_{\max}$. By application, the same equivalence holds between $\mathbb{E}[G(X^n|m)^\rho]$ and $\max_{x^n} P_{X^n|M}(x^n|m)$ for any $m$.

*Remark 3.20* Unlike the repeated guessing setting, Theorem 3.11 holds true even if near-lossless compression is permitted or if a variable rate key is used.

*Remark 3.21* The first step of the encoding for achieving Theorem 3.11 uses negligible secret key to obscure the type of the sequence. This step was not included in the proof of Theorem 3.9. In fact, the purpose of this encoding step in either case is to allow the result to be claimed for all messages $m$ rather than averaged over the distribution of $M$.

## 3.4 Lossy compression, imperfect secrecy

In full generality, secure source coding involves compression that may be lossy and secrecy that may be imperfect. In fact, compression of a source from a continuous distribution or a discrete distribution with infinite entropy must necessarily be lossy, and the theorems of the previous section are not relevant. In this section we outline how the theory is generalized.

New intricacies and trade-offs are revealed when the theory allows for lossy reconstruction by the decoder, measured by average distortion. Under constrained resources, the objectives of source coding can be at odds with each other, forcing the system design to prioritize the quality of the reproduction against the level of security.

### 3.4.1 Rate-distortion theory

In the simplest case, we now have two measures of distortion to consider—the average distortion of the intended reconstruction at the legitimate receiver and the average distortion obtained by an eavesdropper's best reconstruction. The objective is to minimize the former and maximize the latter. However, a useful generalization of these pairwise distortion functions is to consider three-way distortion functions of the form $d(x, \hat{x}, z)$ which depend on the values of the source, legitimate reconstruction, and eavesdropper reconstruction together. We will make use of this generalization in Section 3.5.

As in the lossless compression case, the theory of secure source coding is enriched, and the security made robust, by taking into account causal disclosure. Since the reconstruction is now allowed to be lossy and hence different from the source sequence, we must now consider that both the source and the reconstruction may be obtained causally by an eavesdropper.

**Box 3.5** Causal disclosure

The eavesdropper obtains the past values of the source $X$ and of the reconstruction $\widehat{X}$, or noisy observations of them, before making its reconstruction of the source at the present time. These observations are modeled as outputs $W_x$ and $W_{\hat{x}}$ of two independent memoryless channels $P_{W_x|X}$ and $P_{W_{\hat{x}}|\widehat{X}}$. Full causal disclosure has noise-free channels.

DEFINITION 3.12 Given a memoryless source $\{X_t \sim P_X\}$, causal disclosure channels $P_{W|X}$ and $P_{W_{\hat{x}}|\widehat{X}}$, and a set of three-way distortion functions $\{d_j(x,\hat{x},z)\}_{j=1}^J$, a rate-distortion tuple $(R, R_K, \{D_j\}_{j=1}^J)$ is achievable if there exists a block-length $n$ and an encoder and decoder at rates $R$ and $R_K$ as described in (3.11) and (3.12) such that

$$\min_{\{Z_t = z_t(M, W_x^{t-1}, W_{\hat{x}}^{t-1})\}_t} \mathbb{E}\, d_j(X^n, \widehat{X}^n, Z^n) \geq D_j \quad \forall j = 1, \ldots, J. \tag{3.53}$$

THEOREM 3.13 ([6, Theorem 1]) *The closure of achievable rate-distortion tuples $(R, R_K, \{D_j\}_{j=1}^J)$ for source coding with security robust to causal disclosure is the set of tuples satisfying*

$$\bigcup_{W_x - X - (U,V) - \widehat{X} - W_{\hat{x}}} \left\{ (R, R_K, \{D_j\}_{j=1}^J) : \begin{array}{rl} & R \geq I(X;U,V) \\ & R_K \geq I(W_x, W_{\hat{x}}; V|U) \\ & D_j \leq \min\limits_{z(u)} \mathbb{E}\, d_j(X, \widehat{X}, z(U)) \quad \forall j \end{array} \right\}. \tag{3.54}$$

*Achievability proof sketch* As in the case of lossless compression, two messages are transmitted, and for the first message, a codebook is generated at rate $I(X;U)$ from the $P_U$ distribution to form a covering of the source space. However, the second message uses a superposition code. For each $u^n$ codeword (indexed by the first message) and key value, a codebook is generated at rate $I(X;V|U)$ from the $P_{V|U}$ distribution to form a conditional covering of the source. The decoder uses both messages and the key to identify the $u^n$ and $v^n$ codewords and produces $\widehat{X}^n$ memorylessly and stochastically according to $P_{\widehat{X}|U,V}$.

Again the analysis is assisted by the use of likelihood encoders for both messages. Let us refer to the codeword associated with the first message as $\bar{u}^n$. The second message specifies a mapping of key values to codewords which we will call $\bar{v}^n(k)$. Due to the likelihood encoder, the key value is approximately independent of the two messages, and the posterior distribution of the source and

reconstruction sequences given the key is approximately

$$P_{X^n,\widehat{X}^n|M,K}(x^n,\hat{x}^n|m,k) \approx \prod_{t=1}^{n} P_{X,\widehat{X}|U,V}(x_t,\hat{x}_t|\bar{u}_t,\bar{v}_t(k)) \tag{3.55}$$

The inequality constraint on $R_K$ in (3.54), when made strict and combined with observation (3.55), assures that the uniformly distributed key cannot be decoded even with access to the causal disclosure. Consequently, further analysis shows that this random key causes the marginal distribution at each point in time to be

$$P_{X_t,\widehat{X}_t|M,W_x^{t-1},W_{\hat{x}}^{t-1}}(x,\hat{x}) \approx P_{X,\widehat{X}|U}(x,\hat{x}|\bar{u}_t). \tag{3.56}$$

This produces the claimed distortion levels. □

*Remark 3.22* The most basic application of Theorem 3.13 uses two distortion functions, where $-d_1(x,\hat{x})$ is the distortion of the decoder reconstruction and $d_2(x,z)$ is the distortion of the eavesdropper reconstruction.

*Remark 3.23* A stochastic decoder is crucial for achieving Theorem 3.13. Without it, causal disclosure of the reconstruction can reveal too much about the secret key and allow an eavesdropper to make inferences about future values of the source and reconstruction. This is dual to the wiretap channel setting, where the encoder must be stochastic.

*Remark 3.24* The variable $U$ can be interpreted as a part of the encoding that is not secured due to lack of resources. The variable $V$ represents the remainder of the encoding that is kept fully secure. However, to achieve secrecy of the $V$ component, it is not sufficient to simply apply a one-time-pad to the encoding (or simply transmit the encoding of $V$ through a secure digital channel). The causal disclosure may still reveal the $V$ codeword. Further randomization of the encoding by use of the key is needed.

*Remark 3.25* The optimal average score in a repeated game is solved by Theorem 3.13. Apply the theorem with full causal disclosure and with only one distortion function $d_1(x,\hat{x},z)$, which is the payoff function of the game. The interpretation of Definition 3.12 is that $X^n$ is an i.i.d. state sequence, and a rate-limited communication system with rate-limited secret key is deployed to help one player of the game (with actions $\widehat{X}$) achieve the best possible performance against another player (with actions $Z$) by revealing the state $X$. The causal disclosure accounts for the fact that the players see each other's actions and the state after each play of the game.

*Remark 3.26* The source coding system that achieves Theorem 3.13 actually synthesizes a memoryless broadcast channel according to $P_{\widehat{X},U|X}$ in a very strong sense: An eavesdropper can decode the $U^n$ sequence, and the conditional distribution of $X^n$ and $\widehat{X}^n$ given $U^n$ is precisely that induced by a memoryless broadcast channel, up to a negligible total variation distance.

*Remark 3.27* Theorem 3.13 does not change if (3.53) is changed to

$$\max_{\{Z_t = z_t(M, W_x^{t-1}, W_{\hat{x}}^{t-1})\}_t} \mathbb{P}\left(d_j(X^n, \widehat{X}^n, Z^n) < D_j\right) < \epsilon \quad \forall j = 1, \ldots, J. \tag{3.57}$$

**Example 3.2** *Binary jamming.* Consider an equiprobable binary source $X$, a single distortion function $d(x, \hat{x}, z) = \mathbb{1}\{x = \hat{x} \neq z\}$, and full causal disclosure (i.e. $W_x = X$ and $W_{\hat{x}} = \widehat{X}$). Apply Theorem 3.13 to find the maximum frequency with which $\widehat{X}$ can equal $X$ without being jammed by $Z$, as a function of the compression rate and secret key rate.

Solution

Numerical optimization gives the trade-off depicted in Figure 3.5.

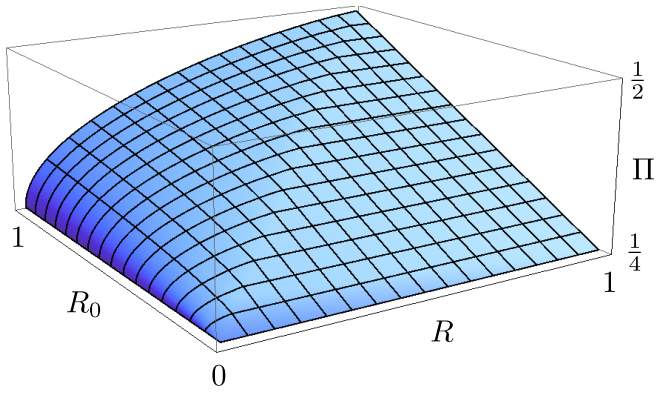


**Figure 3.5** Binary jamming example.

Notice that increases in the key rate $R_K$ continue to benefit the achievable performance even when $R_K > R$. This highlights that a one-time-pad is not a sufficiently general encoder construction for secure source coding.

### 3.4.2 Henchman

We now return to the henchman setting of Figure 3.4 and allow loss in the encoding, measured by average distortion. There are two performance guarantees to certify in this setting. The average distortion of the decoder reconstruction must be less than $D$, and the average distortion of the eavesdropper reconstruction must be greater than $D_E$, even with the assistance of the henchman.

The optimal encoding in the henchman setting is rather simple. For each secret key value, an independent source codebook is used (covering). However, tension between fidelity and security arises in the choice of distribution for which the codebook is designed. It is desirable to encode the source sequence in a way that is not easily refinable by the henchman.

DEFINITION 3.14 Given a memoryless source $\{X_t \sim P_X\}$ and a pair of distortion functions $d(x,\hat{x})$ and $d_E(x,z)$, a rate-distortion tuple $(R, R_K, R_H, D, D_E)$ is achievable if for all $\epsilon > 0$ there exist a blocklength $n$ and an encoder and decoder at compression rate $R$ and key rate $R_K$ as described in (3.11) and (3.12) such that

$$\mathbb{E}\, d(X^n, \widehat{X}^n) \le D, \tag{3.58}$$

$$\min_{\substack{Z^n = z^n(M,M_H),\\ M_H = m_H(X^n,M,K),\\ |\mathcal{M}_H| = 2^{nR_H}}} \mathbb{E}\, d_E(X^n, Z^n) \ge D_E. \tag{3.59}$$

THEOREM 3.15 ([7]) *The closure of achievable rate-distortion tuples $(R, R_K, R_H, D, D_E)$ for source coding with security robust to worst-case side information is the set of tuples satisfying*

$$R \ge I(X;\widehat{X}), \tag{3.60}$$

$$D \ge \mathbb{E}\, d(X,\widehat{X}), \tag{3.61}$$

$$D_E \le \begin{cases} D_E(R_H) & \text{if } R_H < R_K, \\ \min\{D_E(R_H), D_E(R_H - R_K, P_{X,\widehat{X}})\} & \text{if } R_H \ge R_K, \end{cases} \tag{3.62}$$

*for some $P_{X,\widehat{X}} = P_X P_{\widehat{X}|X}$, where $D_E(r)$ is the point-to-point distortion-rate function and $D_E(r, P_{X,\widehat{X}})$ is the point-to-point distortion-rate function with side information channel $P_{\widehat{X}|X}$ to the encoder and decoder:*

$$D_E(r) \triangleq \min_{P_{Z|X}: r \ge I(X;Z)} \mathbb{E}\, d_E(X,Z), \tag{3.63}$$

$$D_E(r, P_{X,\widehat{X}}) \triangleq \min_{P_{Z|X,\widehat{X}}: r \ge I(X;Z|\widehat{X})} \mathbb{E}\, d_E(X,Z). \tag{3.64}$$

*Proof sketch* The converse is again the easier proof. The henchman and the eavesdropper always have the option of compressing the source using a rate-distortion code or revealing the key as part of $M_H$, if $R_H$ is sufficiently high. The non-trivial step in the converse, for $R_H > R_K$, involves conditioning on the joint empirical distribution of the source $X^n$ and the reconstruction $\widehat{X}^n$ to claim the lower bound on the rate $R$ and the upper bound $D_E \le D_E(R_H - R_K, P_{X,\widehat{X}})$. However, the encoding need not produce a consistent empirical distribution for each instantiation of the source sequence. Careful use of empirical distribution bounds imposed by source coding are needed to account for this.

Achievability is proven using an independent codebook construction at rate $R$ for each key value, generated according to the distribution $P_{\widehat{X}}$. A likelihood encoder is used to simplify the secrecy analysis. Analogous to the lossless case, the important step in the analysis is to show that for large $n$ a random sequence $X^n$ that is produced in two steps—first, select a sequence $\widehat{X}^n$ uniformly distributed on a random set (generated according to $P_{\widehat{X}}$), and second, produce $X^n$ memorylessly according to $P_{X|\widehat{X}}$—is with high probability not compressible by the

henchman beyond the trivial schemes associated with explicitly describing the sequence $\widehat{X}^n$ from the codebook or treating $X^n$ as a memoryless sequence. □

*Remark 3.28* If the henchman's rate is below the secret key rate, then the system effectively attains perfect secrecy. The henchman can only describe the source sequence as efficiently as rate-distortion theory allows. On the other hand, if the henchman has enough rate to reveal the secret key then he has a choice to make. He can transmit the key and use any additional rate to refine the source description further, but this will not always be more efficient then describing the source in the absence of the securely encoded message.

*Remark 3.29* If $R_K \geq R$, then Theorem 3.15 simplifies to

$$D \geq D(R), \tag{3.65}$$

$$D_E \leq D_E(R_H). \tag{3.66}$$

This can be achieved by a one-time-pad on the message, although the encoding outlined in the proof achieves this effect in a different way.

*Remark 3.30* Theorem 3.15 does not change if (3.59) is changed to

$$\max_{\substack{Z^n=z^n(M,M_H),\\ M_H=m_H(X^n,M,K),\\ |\mathcal{M}_H|=2^{nR_H}}} \mathbb{P}\left(d_E(X^n,Z^n) < D_E\right) < \epsilon. \tag{3.67}$$

## 3.5 Equivocation

Equivocation, defined in Box 3.6, is the predominant metric for partial secrecy in the information theory literature. Shannon planted this seeds by mentioning equivocation in his secrecy analysis in [2], and it took root when Wyner characterized the optimal equivocation rate for communication through the wiretap channel in [1].

It turns out that the use of equivocation as a secrecy metric is a particular special case of the secure source coding results already discussed in this chapter.

**Box 3.6** Equivocation rate

Equivocation is the conditional entropy of the protected information (usually the source) given everything known to the eavesdropper. Partial secrecy is often characterized by the equivocation rate, which is the equivocation divided by the length of the encoded source sequence. An equivalent quantity is the information leakage rate, which is the mutual information between what the eavesdropper knows and the protected information, again normalized by the blocklength. A high equivocation rate is a low information leakage rate.

In hindsight, we can arrive at the equivocation-rate metric of secrecy by applying Theorem 3.13 (the rate-distortion theorem) with the *log-loss* distortion

function defined in Box 3.7. Applications of the log-loss distortion function to rate-distortion theory were first demonstrated in works such as [11].

**Box 3.7** Logarithmic-loss distortion function

Let $w \in \mathcal{W}$ be discrete valued and $P$ be a distribution on $\mathcal{W}$. The log-loss function $d(w, P)$ is the logarithm of the inverse of the probability of $w$ according to $P$.

$$d(w, P) \triangleq \log \frac{1}{P(w)}. \tag{3.68}$$

The log-loss function is zero if and only if $P$ puts all probability on the outcome $w$.
The optimal choice $P^*$ to minimize the expected log-loss with respect to a random variable $W$ is $P^* = P_W$, and the expected log-loss is $H(W)$. Similarly, if a correlated observation $V = v$ is available, then $P^* = P_{W|V=v}$, and the expected log-loss is $H(W|V)$.

DEFINITION 3.16 Given a memoryless source $\{X_t \sim P_X\}$ and a distortion function $d(x, \hat{x})$, a rate-distortion-equivocation quadruple $(R, R_K, D, E)$ is achievable for *equivocation with respect to the source* if there exists a blocklength $n$ and an encoder and decoder at compression rate $R$ and key rate $R_K$ as described in (3.11) and (3.12) such that

$$\mathbb{E}\, d(X^n, \hat{X}^n) \leq D, \tag{3.69}$$

$$\frac{1}{n} H(X^n|M) \geq E. \tag{3.70}$$

Likewise, $(R, R_K, D, E)$ is achievable for *equivocation with respect to the reconstruction* if (3.70) is replaced by

$$\frac{1}{n} H(\hat{X}^n|M) \geq E, \tag{3.71}$$

and $(R, R_K, D, E)$ is achievable for *equivocation with respect to both the source and reconstruction* if (3.70) is replaced by

$$\frac{1}{n} H(X^n, \hat{X}^n|M) \geq E. \tag{3.72}$$

COROLLARY 3.17 ([6, Corollary 5]) *The closure of achievable rate-distortion-equivocation quadruples $(R, R_K, D, E)$ for each of the three variants of equivocation are as follows:*

*1 For equivocation with respect to the source,*

$$\bigcup_{P_{\hat{X}|X}} \left\{ (R, R_K, D, E) : \begin{array}{c} R \geq I(X;\hat{X}) \\ D \geq \mathbb{E}\, d(X,\hat{X}) \\ E \leq H(X) - [I(X;\hat{X}) - R_K]_+ \end{array} \right\}, \tag{3.73}$$

*where* $[x]_+ = \max\{0, x\}$.

*2 For equivocation with respect to the reconstruction,*

$$\bigcup_{X-U-\hat{X}} \left\{ (R, R_K, D, E) : \begin{array}{c} R \geq I(X;U) \\ D \geq \mathbb{E}\, d(X,\hat{X}) \\ E \leq H(\hat{X}) - [I(\hat{X};U) - R_K]_+ \end{array} \right\}. \tag{3.74}$$

*3 For equivocation with respect to both the source and reconstruction,*

$$\bigcup_{X-U-\hat{X}} \left\{ (R, R_K, D, E) : \begin{array}{c} R \geq I(X;U) \\ D \geq \mathbb{E}\, d(X,\hat{X}) \\ E \leq H(X,\hat{X}) - [I(X,\hat{X};U) - R_K]_+ \end{array} \right\}. \tag{3.75}$$

*Proof sketch* It must first be argued that Theorem 3.13 applies to this equivocation problem. Consider Definition 3.12 with two distortion constraints. For the first distortion constraint set $d_1(x, \hat{x}, z) = -d(x, \hat{x})$ and $D_1 = -D$, where $d(x, \hat{x})$ and $D$ are from Definition 3.16, the equivocation problem statement. The reason for the negative signs is that we intend to minimize this distortion instead of maximizing it.

The choice of the second distortion constraint and the causal disclosure depend on which of the three equivocation claims we wish to prove. In each case, set the second distortion constraint $D_2 = E$. For equivocation with respect to the source, set $d_2(x, \hat{x}, z)$ to be the log-loss function with respect to the source $X_t$ (i.e. set $W = X_t$ in Box 3.7), and let the causal disclosure be full causal disclosure of the source only (i.e. $W_x = X$ and $W_{\hat{x}} = \emptyset$).

Notice what happens to (3.53) with these specifications of log-loss and full causal disclosure of the source.

$$\min_{\{Z_t = z_t(M, W_x^{t-1}, W_{\hat{x}}^{t-1})\}_t} \mathbb{E}\, d_2(X^n, \widehat{X}^n, Z^n) = \min_{\{Z_t = z_t(M, X^{t-1})\}_t} \mathbb{E}\, \frac{1}{n} \sum_{t=1}^{n} \log \frac{1}{Z_t(X_t)} \tag{3.76}$$

$$= \frac{1}{n} \sum_{t=1}^{n} \min_{Z_t = z_t(M, X^{t-1})} \mathbb{E} \log \frac{1}{Z_t(X_t)} \tag{3.77}$$

$$= \frac{1}{n} \sum_{t=1}^{n} H(X_t | M, X^{t-1}) \tag{3.78}$$

$$= \frac{1}{n} H(X^n | M). \tag{3.79}$$

Thus, the second distortion constraint imposed by Definition 3.12 becomes precisely the equivocation constraint of Definition 3.16.

The choices of distortion constraint and causal disclosure for the other equivocation claims are similar. For equivocation with respect to the reconstruction, set $d_2(x, \hat{x}, z)$ to be the log-loss function with respect to the reconstruction, and let the causal disclosure be full causal disclosure of the reconstruction only. For equivocation with respect to both the source and reconstruction, set $d_2(x, \hat{x}, z)$ to be the log-loss function with respect to both the source and reconstruction, and let the causal disclosure be full causal disclosure of both the source and reconstruction. The equivalence between Definition 3.12 and Definition 3.16 follows as above.

What remains is to plug these choices of distortion functions and causal disclosure into Theorem 3.13 and show that the regions simplify to those stated in Corollary 3.17. This is done through simple manipulations that can be found in [6]. □

*Remark 3.31* The causal disclosure present in the problem definition for Theorem 3.13 plays an essential role in the connection to equivocation rate. Without causal disclosure, the chain rule used in (3.79) would be invalid.

*Remark 3.32* Although Statement 1) of Corollary 3.17 is well understood and intuitive, the other two claims were newly introduced to the literature in [6]. Equivocation with respect to something other than the source is not commonly considered. The second two claims are qualitatively different from the first claim in that they use an auxiliary variable, and the accompanying encoding scheme is more complex. They retain some of the complexity of Theorem 3.13, which takes advantage of a stochastic decoder.

*Remark 3.33* While it appears quite difficult to expand the general rate-distortion theory, along the lines of Theorem 3.13, to settings with multiple correlated information sources (e.g. side information at the decoder or separated encoders), the special case of log-loss distortion is much more manageable. Notice how Corollary 3.17 has fewer auxiliary variables than Theorem 3.13. The coding scheme that achieves (3.73) for optimal equivocation with respect to the source is also considerably simpler. In some sense, log-loss distortion (i.e. the study of equivocation rate) allows one to treat partial secrecy as a quantity rather than a quality. The reduced intricacy has yielded tractable solutions in more complex secrecy settings.

## References

[1] A. D. Wyner, "The wire-tap channel," *The Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, Oct 1975.

[2] C. E. Shannon, "Communication theory of secrecy systems," *The Bell System Technical Journal*, vol. 28, no. 4, pp. 656–715, Oct 1949.

[3] Y. Kaspi and N. Merhav, "Zero-delay and causal secure source coding," *IEEE Transactions on Information Theory*, vol. 61, no. 11, pp. 6238–6250, Nov 2015.

[4] C. Uduwerelle, S.-W. Ho, and T. Chan, "Design of error-free perfect secrecy system by prefix codes and partition codes," in *Proc. of IEEE Int'l. Symp. Inf. Theory (ISIT)*, July 2012, pp. 1593–1597.

[5] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, July 1948.

[6] C. Schieler and P. Cuff, "Rate-distortion theory for secrecy systems," *Information Theory, IEEE Transactions on*, vol. 60, no. 12, pp. 7584–7605, Dec 2014.

[7] ——, "The henchman problem: Measuring secrecy by the minimum distortion in a list," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3436–3450, June 2016.

[8] N. Merhav and E. Arikan, "The shannon cipher system with a guessing wiretapper," *IEEE Transactions on Information Theory*, vol. 45, no. 6, pp. 1860–1866, Sep 1999.

[9] E. Arikan, "An inequality on guessing and its application to sequential decoding," *IEEE Transactions on Information Theory*, vol. 42, no. 1, pp. 99–105, Jan 1996.

[10] N. Merhav, "A large-deviations notion of perfect secrecy," *IEEE Transactions on Information Theory*, vol. 49, no. 2, pp. 506–508, Feb 2003.

[11] T. A. Courtade and T. Weissman, "Multiterminal source coding under logarithmic loss," *IEEE Transactions on Information Theory*, vol. 60, no. 1, pp. 740–761, Jan 2014.

# Index