Article

Encoding Individual Source Sequences for the Wiretap Channel

Neri Merhav

The Viterbi Faculty of Electrical and Computer Engineering, Technion-Israel Institute of Technology, Technion City, Haifa 3200003, Israel; merhav@technion.ac.il; Tel.: +972-4-8294737

Abstract: We consider the problem of encoding a deterministic source sequence (i.e., individual sequence) for the degraded wiretap channel by means of an encoder and decoder that can both be implemented as finite-state machines. Our first main result is a necessary condition for both reliable and secure transmission in terms of the given source sequence, the bandwidth expansion factor, the secrecy capacity, the number of states of the encoder and the number of states of the decoder. Equivalently, this necessary condition can be presented as a converse bound (i.e., a lower bound) on the smallest achievable bandwidth expansion factor. The bound is asymptotically achievable by Lempel–Ziv compression followed by good channel coding for the wiretap channel. Given that the lower bound is saturated, we also derive a lower bound on the minimum necessary rate of purely random bits needed for local randomness at the encoder in order to meet the security constraint. This bound too is achieved by the same achievability scheme. Finally, we extend the main results to the case where the legitimate decoder has access to a side information sequence, which is another individual sequence that may be related to the source sequence, and a noisy version of the side information sequence leaks to the wiretapper.

Keywords: wiretap channel; physical layer security; semantic security; individual sequence; finite-state machine; Lempel–Ziv algorithm; side information

1. Introduction

In his seminal paper, Wyner [1] introduced the wiretap channel as a model of secure communication over a degraded broadcast channel, without using a secret key, where the legitimate receiver has access to the output of the good channel and the wiretapper receives the output of the bad channel. The main idea is that the excess noise at the output of the wiretapper channel is utilized to secure the message intended to the legitimate receiver. Wyner fully characterized the best achievable trade-off between reliable communication to the legitimate receiver and the equivocation rate at the wiretapper, which was quantified in terms of the conditional entropy of the source given the output of the wiretapper channel. One of the most important concepts, introduced by Wyner, was the *secrecy capacity*, that is, the supremum of all coding rates that allow both reliable decoding at the legitimate receiver and full secrecy, where the equivocation rate saturates at the (unconditional) entropy rate of the source, or equivalently, the normalized mutual information between the source and the wiretap channel output is vanishingly small for large blocklength. The idea behind the construction of a good code for the wiretap channel is basically the same as the idea of binning: one designs a big code, that can be reliably decoded at the legitimate receiver, which is subdivided into smaller codes that are fed by purely random bits that are unrelated to the secret message. Each such sub-code can be reliably decoded individually by the wiretapper to its full capacity, thus leaving no further decoding capability for the remaining bits, which all belong to the real secret message.

During the nearly five decades that have passed since [1] was published, the wiretap channel model was extended and further developed in many aspects. We mention here just a few. Three years after Wyner, Csiszár and Körner [2] extended the wiretap channel to a general broadcast channel that is not necessarily degraded, allowing also a common

message intended to both receivers. In the same year, Leung-Yan-Cheong and Hellman [3], studied the Gaussian wiretap channel, and proved, among other things, that its secrecy capacity is equal to the difference between the capacity of the legitimate channel and that of the wiretap channel. In [4], Ozarow and Wyner considered a somewhat different model, known as the type II wiretap channel, where the channel to the legitimate receiver is clean (noiseless), and the wiretapper can access a subset of the coded bits. In [5], Yamamoto extended the wiretap channel to include two parallel broadcast channels that connect one encoder and one legitimate decoder, and both channels are wiretapped by wiretappers that do not cooperate with each other. A few years later, the same author [6] further developed the scope of [1] in two ways: first, by allowing a private secret key to be shared between the encoder and the legitimate receiver, and secondly, by allowing a given distortion in the reproducing the source at the legitimate receiver. The main coding theorem of [6] suggests a three-fold separation principle, which asserts that no asymptotic optimality is lost if the encoder first applies a good lossy source code, then encrypts the compressed bits, and finally, applies a good channel code for the wiretap channel. In [7], this model in turn was generalized to allow source side information at the decoder and at the wiretapper in a degraded structure with application to systematic coding for the wiretap channel. The Gaussian wiretap channel model of [3] was also extended in two ways: the first is the Gaussian multiple access wiretap channel of [8], and the second is Gaussian interference wiretap channel of [9,10], where the encoder has access to the interference signal as side information. Wiretap channels with feedback were considered in [11], where it was shown that feedback is best used for the purpose of sharing a secret key as in [6,7]. More recent research efforts were dedicated to strengthening the secrecy metric from weak secrecy to strong secrecy, where the mutual information between the source and the wiretap channel output vanishes, even without normalization by the blocklength, as well as to semantic security, which is similar but refers even to the worst-case message source distribution; see, for example, Refs. [12–16], (Section 3.3 in [14]).

In this work, we look at Wyner's wiretap channel model from a different perspective. Following the individual sequence approach pioneered by Ziv in [17–19], and continued in later works, such as [20,21], we consider the problem of encoding a deterministic source sequence (i.e., an individual sequence) for the degraded wiretap channel using finite-state encoders and finite-state decoders. One of the non-trivial issues associated with individual sequences, in the context of the wiretap channel, is how to define the security metric, as there is no probability distribution assigned to the source, and therefore, the equivocation, or the mutual information between the source and the wiretap channel output, cannot be well defined. In [20], a similar dilemma was encountered in the context of private key encryption of individual sequences, and in the converse theorem therein, it was assumed that the system is perfectly secure in the sense that the probability distribution of the cryptogram does not depend on the source sequence. In principle, it is possible to apply the same approach here, where the word 'cryptogram' is replaced by the 'wiretap channel output'. However, in order to handle residual dependencies, which will always exist, it would be better to use a security metric that quantifies those small dependencies. To this end, it makes sense to adopt the above-mentioned maximum mutual information security metric (or, equivalently the semantic security metric), where the maximum is over all input assignments. After this maximization, this quantity depends only on the 'channel' between the source and the wiretap channel output.

Our first main result is a necessary condition (i.e., a converse to a coding theorem) for both reliable and secure transmission, which depends on: (i) the given individual source sequence, (ii) the bandwidth expansion factor, (iii) the secrecy capacity, (iv) the number of states of the encoder, (v) the number of states of the decoder, (vi) the allowed bit error probability at the legitimate decoder and (vii) the allowed maximum mutual information secrecy. Equivalently, this necessary condition can be presented as a converse bound (i.e., a lower bound) to the smallest achievable bandwidth expansion factor. The bound is asymptotically achievable by Lempel–Ziv (LZ) compression followed by a good channel coding scheme for the wiretap channel. Given that this lower bound is saturated, we then derive also a lower bound on the minimum necessary rate of purely random bits needed for adequate local randomness at the encoder, in order to meet the security constraint. This bound too is achieved by the same achievability scheme, a fact which may be of independent interest regardless of individual sequences and finite-state encoders and decoders (i.e., also for ordinary block codes in the traditional probabilistic setting). Finally, we extend the main results to the case where the legitimate decoder has access to a side information sequence, which is another individual sequence that may be related to the source sequence, and where a noisy version of the side information sequence leaks to the wiretapper. It turns out that in this case, the best strategy is the same as if one assumes that the wiretapper sees the clean side information sequence. While this may not be surprising as far as sufficiency is concerned (i.e., as an achievability result), it is less obvious in the context of necessity (i.e., a converse theorem).

The remaining part of this article is organized as follows. In Section 2, we establish the notation, provide some definitions and formalize the problem setting. In Section 3, we provide the main results of this article and discuss them in detail. In Section 4, the extension that incorporates side information is presented. Finally, in Section 5, the proofs of the main theorems are given.

2. Notation, Definitions, and Problem Setting

2.1. Notation

Throughout this paper, random variables are denoted by capital letters; specific values they may take are denoted by the corresponding lower case letters; and their alphabets are denoted by calligraphic letters. Random vectors, their realizations, and their alphabets are denoted, respectively, by capital letters, the corresponding lower case letters and calligraphic letters, all superscripted by their dimensions. For example, the random vector $X^n = (X_1, \ldots, X_n)$, (n – positive integer) may take a specific vector value $x^n = (x_1, \ldots, x_n)$ in $\mathcal{X}^n$, the n-th order Cartesian power of $\mathcal{X}$, which is the alphabet of each component of this vector. Infinite sequences are denoted using the bold face font, e.g., $\boldsymbol{x} = (x_1, x_2, \ldots)$. Segments of vectors are denoted by subscripts and superscripts that correspond to the start and the end locations; for example, x_i^j, for $i < j$ integers, denotes $(x_i, x_{i+1}, \ldots, x_j)$. When $i = 1$, the subscript is omitted.

Sources and channels are denoted by the letter P or Q, subscripted by the names of the relevant random variables/vectors and their conditionings, if applicable, following the standard notation conventions, e.g., Q_X, $P_{Y|X}$, and so on, or by abbreviated names that describe their functionality. When there is no room for ambiguity, these subscripts are omitted. The probability of an event $\mathcal{E}$ will be denoted by $\Pr\{\mathcal{E}\}$, and the expectation operator with respect to (w.r.t.) a probability distribution P is denoted by $\mathbb{E}_P\{\cdot\}$. Again, the subscript is omitted if the underlying probability distribution is clear from the context or explicitly explained in the following text. The indicator function of an event $\mathcal{E}$ is denoted by $1\{\mathcal{E}\}$, that is, $1\{\mathcal{E}\} = 1$ if $\mathcal{E}$ occurs; otherwise, $1\{\mathcal{E}\} = 0$.

Throughout considerably large parts of the paper, the analysis is carried out w.r.t. joint distributions that involve several random variables. Some of these random variables are induced from empirical distributions of deterministic sequences, while others are ordinary random variables. Random variables from the former kind are denoted with 'hats'. As a simple example, consider a deterministic sequence, x^n, that is fed as an input to a memoryless channel defined by a single-letter transition matrix, $\{P_{Y|X}, x \in \mathcal{X}, y \in \mathcal{Y}\}$, and let y^n denote a realization of the corresponding channel output. Let $P_{\hat{X}\hat{Y}}(x,y) = \frac{1}{n}\sum_{i=1}^{n} 1\{x_i = x, y_i = y\}$ denote the joint empirical distribution induced from (x^n, y^n). In addition to $P_{\hat{X}\hat{Y}}(x,y)$, we also define $P_{\hat{X}Y}(x,y) = \mathbb{E}\{P_{\hat{X}\hat{Y}}(x,y)\}$, where now Y is an ordinary random variable. Clearly, the relation between the two distributions is given by $P_{\hat{X}Y}(x,y) = P_{\hat{X}}(x) \cdot P_{Y|X}(y|x)$, where $P_{\hat{X}}(x) = \sum_{y} P_{\hat{X}\hat{Y}}(x,y)$ is the empirical marginal of $\hat{X}$. Such mixed joint distributions underlie certain information-theoretic quantities, for example, $I(\hat{X}; Y)$ and $H(Y|\hat{X})$ denote, respectively, the mutual information between $\hat{X}$ and Y and the conditional entropy of Y given $\hat{X}$, both induced from $P_{\hat{X}Y}$. The same notation rules are applicable in more involved situations too.

2.2. Definitions and Problem Setting

We consider the system configuration of the degraded wiretap channel, depicted in Figure 1. Let $\boldsymbol{u} = (u_1, u_2, \ldots)$ be a deterministic source sequence (i.e., individual sequence), whose symbols take values in a finite alphabet, $\mathcal{U}$, of size α. This source sequence is to be conveyed reliably to a legitimate decoder while keeping it secret from a wiretapper, as described below. The encoding mechanism is as follows. The source sequence, $\boldsymbol{u}$ is first is divided into chunks of length k, $\tilde{u}_i = u_{ik+1}^{ik+k} \in \mathcal{U}^k$, $i = 0, 1, 2, \ldots$, which are fed into a stochastic finite-state encoder, defined by the following equations:

$$\Pr\{\tilde{X}_i = \tilde{x}|\tilde{u}_i = \tilde{u}, s_i^{\mathrm{e}} = s\} = P(\tilde{x}|\tilde{u}, s), \quad i = 0, 1, 2, \ldots \tag{1}$$

$$s_{i+1}^{\mathrm{e}} = h(\tilde{u}_i, s_i^{\mathrm{e}}), \quad i = 0, 1, 2, \ldots. \tag{2}$$

We allow a stochastic encoder, in view of the fact that even the traditional, probabilistic setting, optimal coding for the wiretap channel (see [1] and later works) must be randomized in order to meet the security requirements. Here, $\tilde{X}_i$ is a random vector taking values in $\mathcal{X}^m$, $\mathcal{X}$ being the β-ary input alphabet of the channel and m being a positive integer; $\tilde{x} \in \mathcal{X}^m$ is a realization of $\tilde{X}_i$; and s_i^{e} is the state of the encoder at time i, which designates the memory of the encoder with regard to the past of the source sequence. In other words, at time instant i, whatever the encoder 'remembers' from $(u_1, u_2, \ldots, u_{i-1})$ is stored in the variable s_i^{e} (for example, in the case of trellis coding, it can be a shift register of finite length, say p, that stores the most recent symbols of p, $(u_{i-p}, u_{i-p+1}, \ldots, u_{i-1})$, or, in block coding, it can be the contents of the current block starting from its beginning up to the present). The state, s_i^{e}, takes values in a finite set of states, $\mathcal{S}^{\mathrm{e}}$, of size q_{e}. In the above equation, the variable $\tilde{u}$ is any member of $\mathcal{U}^k$. The function $h : \mathcal{U}^k \times \mathcal{S}^{\mathrm{e}} \to \mathcal{S}^{\mathrm{e}}$ is called the *next-state function* of the encoder. (More generally, we could define both s_{i+1}^{e} and $\tilde{x}_i$ to be random functions of the $(\tilde{u}_i, s_i^{\mathrm{e}})$ by a conditional joint distribution, $\Pr\{\tilde{X}_i = \tilde{x}, s_{i+1}^{\mathrm{e}} = s|\tilde{u}_i = \tilde{u}, s_i^{\mathrm{e}} = s'\}$. However, it makes sense to let the encoder state sequence evolve deterministically in response to the input $\boldsymbol{u}$ since the state designates the memory of the encoder to past inputs.) Finally, $P(\tilde{x}|\tilde{u}, s)$, $\tilde{u} \in \mathcal{U}^k$, $s \in \mathcal{S}^{\mathrm{e}}$, $\tilde{x} \in \mathcal{X}^m$, is a conditional probability distribution function, i.e., $\{P(\tilde{x}|\tilde{u}, s)\}$ are all non-negative and $\sum_{\tilde{x}} P(\tilde{x}|\tilde{u}, s) = 1$ for all $(\tilde{u}, s) \in \mathcal{U}^k \times \mathcal{S}^{\mathrm{e}}$. The vector $\tilde{x}_i$ designates the current output vector from the encoder in response to the current input source vector, $\tilde{u}_i$ and its current state, s_i^{e}. Without loss of generality, we assume that the initial state of the encoder, s_0^{e}, is some fixed member of $\mathcal{S}^{\mathrm{e}}$. The ratio

$$\lambda \triangleq \frac{m}{k} \tag{3}$$

is referred to as the *bandwidth expansion factor*. It should be pointed out that the parameters k and m are fixed integers, which are not necessarily large (e.g., $k = 2$ and $m = 3$ are valid values of k and m). The concatenation of the output vectors from the encoder, $\tilde{x}_0, \tilde{x}_1, \ldots,$ is viewed as a sequence chunks of channel input symbols, $x_1, x_2, \ldots,$ with $\tilde{x}_i = x_{im+1}^{im+m},$ similarly to the above-defined partition of the source sequence.

The sequence of encoder outputs, $x_1, x_2, \ldots,$ is fed into a discrete memoryless channel (DMC), henceforth referred to as the *main channel*, whose corresponding outputs, $y_1, y_2, \ldots,$ are generated according to

$$\Pr\{Y^N = y^N | X^N = x^N\} = Q_{\mathrm{M}}(y^N | x^N) = \prod_{i=1}^{N} Q_{\mathrm{M}}(y_i | x_i), \tag{4}$$

for every positive integer N and every $x^N \in \mathcal{X}^N$ and $y^N \in \mathcal{Y}^N$. The channel output symbols, $\{y_i\}$, take values in a finite alphabet, $\mathcal{Y}$, of size γ.

The sequence of channel outputs, $y_1, y_2, \ldots,$ is divided into chunks of length m, $\tilde{y}_i = y_{im+1}^{im+m}$, $i = 0, 1, 2, \ldots,$ which are fed into a deterministic finite-state decoder, defined according to the following recursive equations:

$$\tilde{v}_i = f(\tilde{y}_i, s_i^{\mathrm{d}}) \tag{5}$$
$$s_{i+1}^{\mathrm{d}} = g(\tilde{y}_i, s_i^{\mathrm{d}}), \tag{6}$$

where the variables in the equations are defined as follows: $\{s_i^{\mathrm{d}}\}$ is the sequence of states of the decoder (which again, designate the finite memory, this time, at the decoder). Each s_i^{d} takes values in a finite set, $\mathcal{S}^{\mathrm{d}}$ of size q_{d}. The variable $\tilde{v}_i \in \mathcal{U}^k$ is the i-th chunk of k source reconstruction symbols, i.e., $\tilde{v}_i = v_{ik+1}^{ik+k}$, $i = 0, 1, \ldots,$ which form the decoder output. The function $f : \mathcal{Y}^m \times \mathcal{S}^{\mathrm{d}} \to \mathcal{U}^k$ is called the *output function* of the decoder and the function $g : \mathcal{Y}^m \times \mathcal{S}^{\mathrm{d}} \to \mathcal{S}^{\mathrm{d}}$ is the next-state function of the decoder. The concatenation of the decoder output vectors, $\tilde{v}_0, \tilde{v}_1, \ldots,$ forms the entire stream of reconstruction symbols, $v_1, v_2, \ldots.$

The output of the main channel, $y_1, y_2, \ldots,$ is fed into another DMC, henceforth referred to as the *wiretap channel*, which generates, in response, a corresponding sequence, $z_1, z_2, \ldots,$ according to

$$\Pr\{Z^N = z^N | Y^N = y^N\} = Q_{\mathrm{w}}(z^N | y^N) = \prod_{i=1}^{N} Q_{\mathrm{w}}(z_i | y_i), \tag{7}$$

where $\{Z_i\}$ and $\{z_i\}$ take values in a finite alphabet $\mathcal{Z}$. We denote the cascade of channels Q_M and Q_W by Q_{MW}, that is

$$Q_{\mathrm{MW}}(z|x) = \sum_{y \in \mathcal{Y}} Q_{\mathrm{M}}(y|x) Q_{\mathrm{w}}(z|y). \tag{8}$$

We seek a communication system (P, h, f, g) which satisfies two requirements:

1. For a given $\epsilon_{\mathrm{r}} > 0$, the system satisfies the following reliability requirement: The bit error probability is guaranteed to be less than ϵ_{r}, i.e.,

$$P_{\mathrm{b}} \overset{\triangle}{=} \frac{1}{k} \sum_{i=1}^{k} \Pr\{V_i \neq u_i\} \leq \epsilon_{\mathrm{r}} \tag{9}$$

for every $(u_1, \ldots, u_k)$ and every combination of initial states of the encoder and the decoder, where $\Pr\{\cdot\}$ is defined w.r.t. the randomness of the encoder and the main channel.

2. For a given $\epsilon_s > 0$, the system satisfies the following security requirement: For every sufficiently large positive integer n,

$$\max_{\mu} I_{\mu}(U^n; Z^N) \leq n\epsilon_s, \tag{10}$$

where $N = n\lambda$ and $I_{\mu}(U^n; Z^N)$ is the mutual information between U^n and Z^N, induced by an input distribution $\mu = \{\mu(u^n),\ u^n \in \mathcal{U}^n\}$ and the system, $\{P(z^N | u^n),\ u^n \in \mathcal{U}^n, z^N \in \mathcal{Z}^N\}$.

As for the reliability requirement, note that the larger k is, the less stringent the requirement becomes. Concerning the security requirement, ideally, we would like to have perfect secrecy, which means that $P(z^N | u^n)$ would be independent of u^n (see also [20]), but it is more realistic to allow a small deviation from this idealization. This security metric is actually the maximum mutual information metric, or equivalently (see [15]) the semantic security, as mentioned in the Introduction.

2.3. Preliminaries and Background

We need two more definitions along with some background associated with them. The first is the *secrecy capacity* [1,14], which is the supremum of all coding rates for which there exist block codes that maintain both an arbitrarily small error probability at the legitimate decoder and an equivocation arbitrarily close to the unconditional entropy of the source. The secrecy capacity is given by

$$C_s = \max_{P_X} I(X; Y | Z) = \max_{P_X} [I(X; Y) - I(X; Z)], \tag{11}$$

with $P_{XYZ}(x, y, z) = P_X(x) \times Q_M(y|x) Q_W(z|y)$ for all $(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$.

The second quantity we need to define is the LZ complexity [22]. In their famous paper [22], Ziv and Lempel actually developed a deterministic counterpart of source coding theory, where instead of imposing assumptions on probabilistic mechanisms that generate the data (i.e., memoryless sources, Markov sources, and general stationary sources), whose relevance to real-world data compression may be subject to dispute, they considered arbitrary, deterministic source sequences (i.e., individual sequences, in their terminology), but they imposed instead a limitation on the resources of the encoder (or the data compression algorithm): they assumed that it has limited storage capability (i.e., limited memory) of past data when encoding the current source symbol. This limited storage was modeled in terms of a finite-state machine, where the state variable of the encoder evolves recursively in time in response to the input and designates the information that the encoder 'remembers' from the past input (just like in the model description in Section 2.2 above). As mentioned earlier, a simple example of such a state variable can be the contents of a finite shift register, fed sequentially by the source sequence in which case the state contains a finite number of the most recent source symbols. This individual-sequence approach is appealing, because it is much more realistic to assume practical limitations on the encoder (which is under the control of the system designer) than to make assumptions on the statistics of the data to be compressed.

Ziv and Lempel developed an asymptotically optimal, practical compression algorithm (which is used in almost every computer), that is well known as the Lempel–Ziv (LZ) algorithm. This algorithm has several variants. One of them, which is called the LZ78 algorithm (where '78' designates the year 1978), is based on the notion of *incremental parsing*: given the the source vector, u^n, the incremental parsing procedure sequentially parses this sequence into distinct phrases such that each new parsed phrase is the shortest string that has not been obtained before as a phrase, with a possible exception of the last phrase, which might be incomplete. Let $c(u^n)$ denote the number of resulting phrases. For example, if $n = 10$ and $u^{10} = (0000110110)$, then incremental parsing (from left to right) yields $(0, 00, 01, 1, 011, 0)$ and so, $c(u^{10}) = 6$. We define the *LZ complexity* of the individual sequence, u^n, as

$$\rho_{\mathrm{LZ}}(u^n) \triangleq \frac{c(u^n) \log c(u^n)}{n}. \tag{12}$$

As was shown by Ziv and Lempel in their seminal paper [22], for large n, the LZ complexity, $\rho_{\mathrm{LZ}}(u^n)$, is essentially the best compression ratio that can be achieved by any information lossless, finite-state encoder (up to some negligibly small terms, for large n), and it can be viewed as the individual-sequence analogue of the entropy rate.
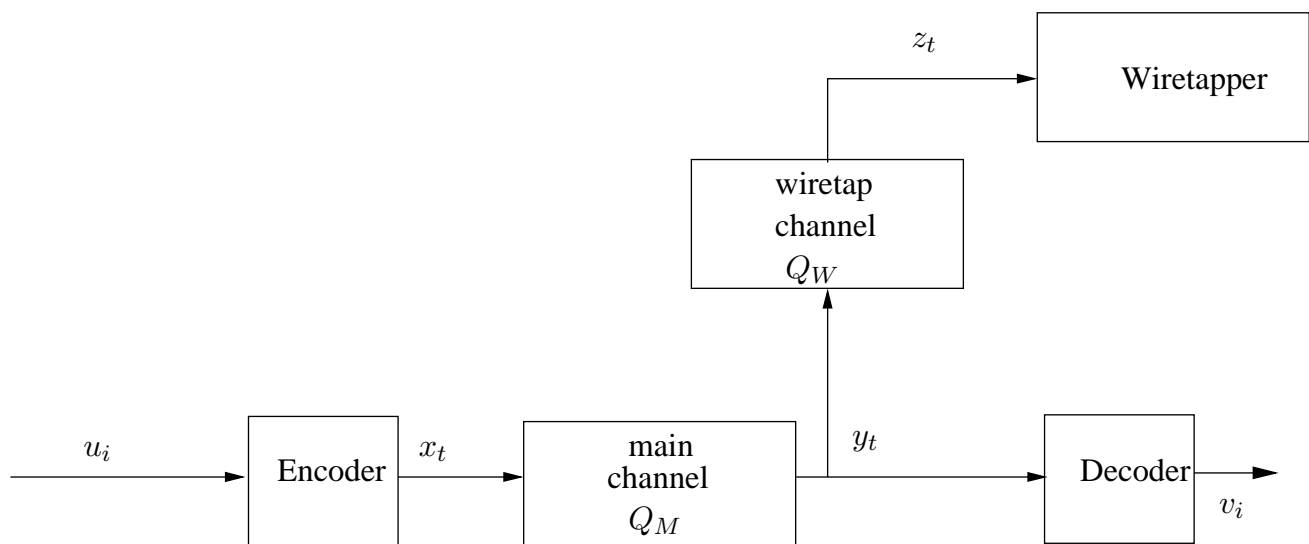
Figure 1. Wiretap channel model. Since the source and the channel may operate at different rates (λ channel symbols per source symbol), the time variables associated with source-related sequences and channel-related sequences are denoted differently, i.e., i and t, respectively.

3. Results

Before moving on to present our first main result, a simple comment is in order. Even in the traditional probabilistic setting, given a source with entropy H and a channel with capacity C, reliable communication cannot be accomplished unless $H \leq \lambda C$, where λ is the bandwidth expansion factor. Since both H and C are given and only λ is under the control of the system designer, it is natural to state this condition as a lower bound to bandwidth expansion factor, i.e., $\lambda \geq H/C$. By the same token, in the presence of a secrecy constraint, λ must not fall below H/C_s. Our converse theorems for individual sequences are presented in the same spirit, where the entropy H at the numerator is replaced by an expression whose main term is the Lempel–Ziv compressibility.

We assume, without essential loss of generality, that k divides n (otherwise, omit the last $(n \bmod k)$ symbols of u^n and replace n by $k \cdot \lfloor n/k \rfloor$ without affecting the asymptotic behavior as $n \to \infty$). Our first main result is the following.

Theorem 1. *Consider the problem setting defined in Section 2. If there exists a stochastic encoder with q_e states and a decoder with q_d states that together satisfy the reliability constraint (9) and the security constraint (10), then the bandwidth expansion factor λ must be lower bounded as follows.*

$$\lambda \geq \frac{\rho_{\mathrm{LZ}}(u^n) - \Delta(\epsilon_r) - \epsilon_s - \zeta_n(q_d, k)}{C_s}, \tag{13}$$

where

$$\Delta(\epsilon_r) \triangleq h_2(\epsilon_r) + \epsilon_r \cdot \log(\alpha - 1), \tag{14}$$

with $h_2(\epsilon_r) = -\epsilon_r \log \epsilon_r - (1 - \epsilon_r) \log(1 - \epsilon_r)$ being the binary entropy function, and

$$\zeta_n(q_d, k) = \min_{\{\ell \text{ divides } n/k\}} \left[\frac{\log q_d + 1}{k\ell} + \frac{2k\ell(\log \alpha + 1)^2}{(1 - \epsilon_n) \log n} + \frac{2k\ell\alpha^{2k\ell} \log \alpha}{n} \right], \tag{15}$$

with $\epsilon_n \to 0$ as $n \to \infty$.

The proof of Theorem 1, like all other proofs in this article, is deferred to Section 5.
Discussion. *A few comments are in order with regard to Theorem 1.*

1. *Irrelevance of q_e.* It is interesting to note that as far as the encoding and decoding resources are concerned, the lower bound depends on k and q_d, but not on the number of states of the encoder, q_e. This means that the same lower bound continues to hold, even if the encoder has an unlimited number of states. Pushing this to the extreme, even if the encoder has room to store the entire past, the lower bound of Theorem 1 would remain unaltered. The crucial bottleneck is, therefore, in the finite memory resources associated with the decoder, where the memory may help to reconstruct the source by exploiting empirical dependencies with the past. The dependence on q_e, however, appear later when we discuss local randomness resources as well as in the extension to the case of decoder side information.

2. *The redundancy term $\zeta_n(q_d, k)$.* A technical comment is in order concerning the term $\zeta_n(q_d, k)$, which involves minimization over all divisors of n/k, where we have already assumed that n/k is integer. Strictly speaking, if n/k happens to be a prime, this minimization is not very meaningful, as $\zeta_n(q_d, k)$ would be relatively large. If this the case, a better bound is obtained if one omits some of the last symbols of u^n, thereby reducing n to, say, n' so that n'/k has a richer set of factors. Consider, for example, the choice $\ell = \ell_n = \lfloor \sqrt{\log n} \rfloor$ (instead of minimizing over ℓ) and replace n/k by the $n/k - (n/k \bmod \ell_n)$, without essential loss of tightness. This way, $\zeta_n(q_d, k)$ would tend to zero as $n \to \infty$, for fixed k and q_d.

3. *Achievability.* Having established that $\zeta_n(q_d, k) \to 0$, and given that ϵ_r and ϵ_s are small, it is clear that the main term at the numerator of the lower bound of Theorem 1 is the term $\rho_{\mathrm{LZ}}(u^n)$, which is, as mentioned earlier, the individual-sequence analogue of the entropy of the source [22]. In other words, λ cannot be much smaller than $\lambda_{\mathrm{L}}(u^n) = \rho_{\mathrm{LZ}}(u^n)/C_s$. A matching achievability scheme would most naturally be based on separation: first apply variable-rate compression of u^n to about $n\rho_{\mathrm{LZ}}(u^n)$ bits using the LZ algorithm [22], and then feed the resulting compressed bit-stream into a good code for the wiretap channel [1] with codewords of length about

$$N = n\lambda_{\mathrm{L}}(u^n) \sim \frac{n\rho_{\mathrm{LZ}}(u^n)}{C_s(1 - \delta)}, \tag{16}$$

where δ is an arbitrarily small (but positive) margin to keep the coding rate strictly smaller than C_s. However, to this end, the decoder must know N. One possible solution is that before the actual encoding of each u^n, one would use a separate, auxiliary fixed code that encodes the value of the number of compressed bits, $n\rho_{LZ}(u^n)$, using $\log(n \log \alpha)$ bits (as $n \log \alpha$ is about the number of possible values that $n\rho_{LZ}(u^n)$ can take) and protect it using a channel code of rate less than $C_s(1 - \delta)$. Since the length of this auxiliary code grows only logarithmically with n (as opposed to the 'linear' growth of $n\rho_{LZ}(u^n)$), the overhead in using the auxiliary code is asymptotically negligible. The auxiliary code and the main code are used alternately: first the auxiliary code, and then the main code for each n-tuple of the source. The main channel code is actually an array of codes, one for each possible value of $n\rho_{LZ}(u^n)$. Once the auxiliary decoder has decoded this number, the corresponding main decoder is used. Overall, the resulting bandwidth expansion factor is about

$$\lambda \approx \frac{n\rho_{LZ}(u^n) + \log(n \log \alpha)}{nC_s(1 - \delta)} = \frac{\rho_{LZ}(u^n)}{C_s(1 - \delta)} + O\left(\frac{\log n}{n}\right). \tag{17}$$

Another, perhaps simpler and better, approach is to use the LZ algorithm in the mode of a variable-to-fixed length code: let the length of the channel codeword, N, be fixed, and start to compress $\boldsymbol{u} = (u_1, u_2, \ldots)$ until obtaining $n\rho_{LZ}(u^n) = N \cdot C_s(1 - \delta)$ compressed bits. Then,

$$\lambda = \frac{N}{n} = \frac{\rho_{LZ}(u^n)}{C_s(1 - \delta)}. \tag{18}$$

Of course, these coding schemes require decoder memory that grows exponentially in n, and not just a fixed number, q_d, and therefore strictly speaking, there is a gap between the achievability and the converse result of Theorem 2. However, this gap is closed asymptotically, once we take the limit of $q_d \to 0$ after the limit $n \to \infty$, and we consider successive application of these codes over many blocks. The same approach appears also in [17–19,22] as well as in later related work.

This concludes the discussion on Theorem 1. □

We next focus on local randomness resources that are necessary when the full secrecy capacity is exploited. Specifically, suppose that the stochastic encoder $\{P(\tilde{x}|\tilde{u}, s), \tilde{x} \in \mathcal{X}^n, \tilde{u} \in \mathcal{U}^k, s \in \mathcal{S}^e\}$ is implemented as a deterministic encoder with an additional input of purely random bits, i.e.,

$$\tilde{x}_i = a(\tilde{u}_i, s_i^e, \tilde{b}_i), \tag{19}$$

where $\tilde{b}_i = b_{ij+1}^{ij+j}$ is a string of j purely random bits. The question is the following: how large must j be in order to achieve full secrecy? Equivalently, what is the minimum necessary rate of random bits for local randomness at the encoder for secure coding at the maximum reliable rate? In fact, this question may be interesting on its own right, regardless of the individual-sequence setting and finite-state encoders and decoders, but even for ordinary block coding (which is the special case of $q_e = q_d = 1$) and in the traditional probabilistic setting. The following theorem answers this question.

Theorem 2. *Consider the problem setting defined in Section 2 and let λ meet the lower bound of Theorem 1. If there exists an encoder (19) with q_e states and a decoder with q_d states that jointly satisfy the reliability constraint (9) and the security constraint (10), then*

$$j \geq mI(X^*; Z^*) - k\epsilon_s - \frac{\log q_e}{\ell} \tag{20}$$

where X^* is the random variable that achieves C_s and ℓ is the achiever of $\zeta_n(q_d, k)$.

Note that the lower bound of Theorem 2 depends on $q_{\rm e}$, as opposed to Theorem 1, where it depends only on $q_{\rm d}$. Since $\epsilon_{\rm s}$ is assumed small and $\ell \to \infty$, it is clear that main term is $mI(X^*; Z^*)$, i.e., the bit rate must be essentially at least as large as $I(X^*; Z^*)$ random bits per channel use, or equivalently, $\lambda I(X^*; Z^*)$ bits per source symbol. It is interesting to note that Wyner's code [1] asymptotically achieves this bound when the coding rate saturates the secrecy capacity because the subcode that can be decoded by the wiretapper (within each given bin) is of the rate of about $I(X^*; Z^*)$, and it encodes just the bits of the local randomness. So when working at the full secrecy capacity, Wyner's code is optimal not only in terms of the optimal trade-off between reliability and security, but also in terms of minimum consumption of local, purely random bits.

4. Side Information at the Decoder with Partial Leakage to the Wiretapper

Consider next an extension of our model to the case where there are side information sequences, $w^n = (w_1, \ldots, w_n)$ and $\dot{w}^n = (\dot{w}_1, \ldots, \dot{w}_n)$, available to the decoder and the wiretapper, respectively; see Figure 2. For the purpose of a converse theorem, we assume that w^n is available to the encoder too, whereas in the achievability part, we comment also on the case where it is not. We assume that w^n is a deterministic sequence, but $\dot{w}^n$ is a realization of a random vector $\dot{W}^n = (\dot{W}_1, \ldots, \dot{W}_n)$, which is a noisy version of w^n. In other words, it is generated from w^n by another memoryless channel, $Q_{\dot{W}^n|W^n}(\dot{w}^n|w^n) = \prod_{i=1}^{n} Q_{\dot{W}|W}(\dot{w}_i|w_i)$. The symbols of $\{w_i\}$ and $\{\dot{w}_i\}$ take values in finite alphabets, $\mathcal{W}$ and $\dot{\mathcal{W}}$, respectively. There are two extreme important special cases: (i) $\dot{W}^n = w^n$ almost surely, which is the case of totally insecure side information that fully leaks to the wiretapper, and (ii) $\dot{W}^n$ is degenerated (or independent of w^n), which is the case of secure side information with no leakage to the wiretapper. Every intermediate situation between these two extremes is a situation of partial leakage. The finite-state encoder model is now re-defined according to

$$\Pr\{\tilde{X}_i = \tilde{x} | \tilde{u}_i = \tilde{u}, \ \tilde{w}_i = \tilde{w}, s_i^{\rm e} = s\} \quad = \quad P(\tilde{x}|\tilde{u}, \tilde{w}, s), \quad i = 0, 1, 2, \ldots \tag{21}$$

$$s_{i+1}^{\rm e} \quad = \quad h(\tilde{u}_i, \tilde{w}_i, s_i^{\rm e}), \quad i = 0, 1, 2, \ldots, \tag{22}$$

where $\tilde{w}_i = w_{ik+1}^{ik+k}$, $i = 0, 1, \ldots, n/k - 1$. Likewise, the decoder is given by

$$\tilde{v}_i \quad = \quad f(\tilde{y}_i, \tilde{w}_i, s_i^{\rm d}) \tag{23}$$

$$s_{i+1}^{\rm d} \quad = \quad g(\tilde{y}_i, \tilde{w}_i, s_i^{\rm d}), \tag{24}$$

and the wiretapper has access to Z^N and $\dot{W}^n$. Accordingly, the security constraint is modified as follows: for a given $\epsilon_{\rm s} > 0$ and for every sufficiently large n,

$$\max_{\mu} I_{\mu}(U^n; Z^N | \dot{W}^n) \leq n\epsilon_{\rm s}, \tag{25}$$

where $I_{\mu}(U^n; Z^N | \dot{W}^n)$ is the conditional mutual information between U^n and Z^N given $\dot{W}^n$, induced by $\mu = \{\mu(u^n, \dot{w}^n), \ u^n \in \mathcal{U}^n, \ \dot{w}^n \in \dot{\mathcal{W}}^n\}$ and the system, $\{P(z^N|u^n), \ u^n \in \mathcal{U}^n, z^N \in \mathcal{Z}^N\}$, where $\mu(u^n, \dot{w}^n) = \sum_{w^n} \mu(u^n, w^n) Q_{\dot{W}^n|W^n}(\dot{w}^n|w^n)$.
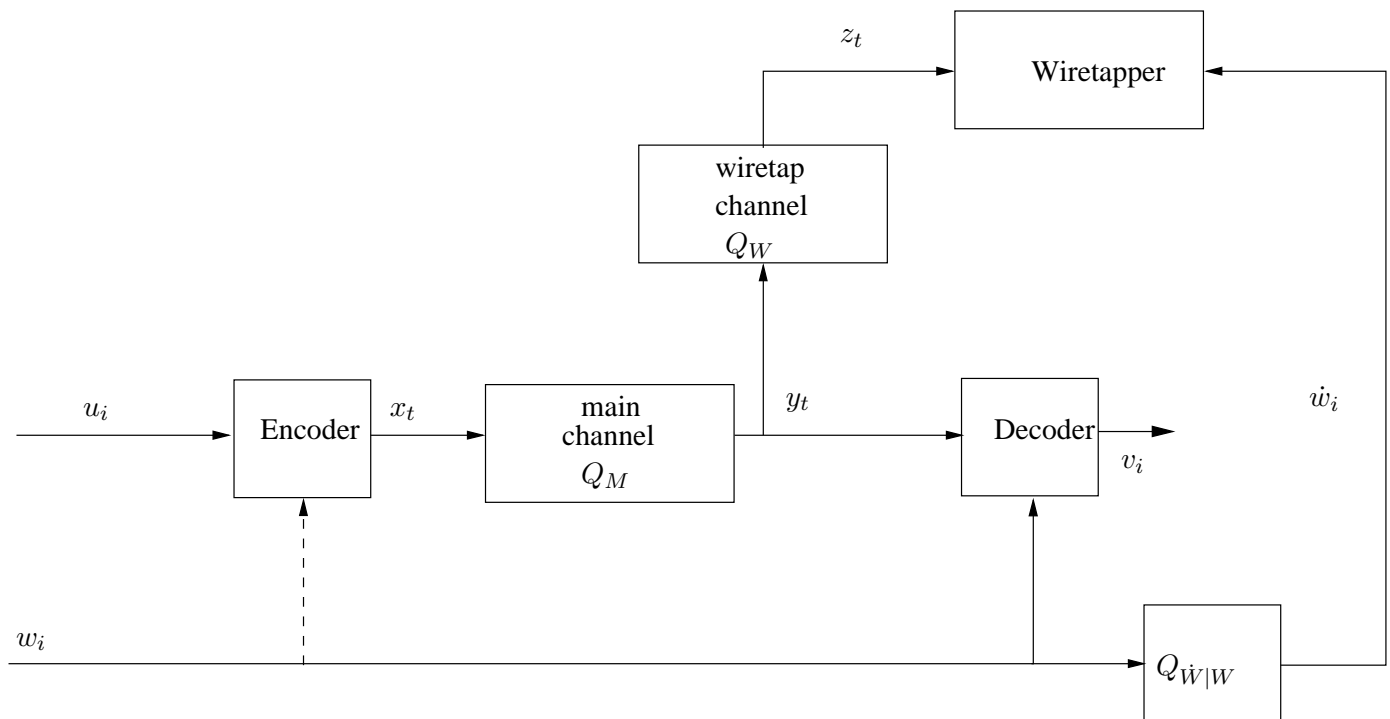
Figure 2. Wiretap channel model with side information.

In order to present the extension of Theorem 1 to incorporate side information, we first need to define the extension of the LZ complexity to include side information, namely, to define the conditional LZ complexity (see also [23]). Given u^n and w^n, let us apply the incremental parsing procedure of the LZ algorithm to the sequence of pairs $((u_1, w_1), (u_2, w_2), \ldots, (u_n, w_n))$. According to this procedure, all phrases are distinct with a possible exception of the last phrase, which might be incomplete. Let $c(u^n, w^n)$ denote the number of distinct phrases. As an example (which appears also in [23]), if

$$
\begin{aligned}
u^6 &= 0\,|\,1\,|\,0\,0\,|\,0\,1| \\
w^6 &= 0\,|\,1\,|\,0\,1\,|\,0\,1|
\end{aligned}
$$

then $c(u^6, w^6) = 4$. Let $c(w^n)$ denote the resulting number of distinct phrases of w^n, and let $w(l)$ denote the l-th distinct w-phrase, $l = 1, 2, ..., c(w^n)$. In the above example, $c(w^6) = 3$. Denote by $c_l(u^n|w^n)$ the number of occurrences of $w(l)$ in the parsing of w^n, or equivalently, the number of distinct u-phrases that jointly appear with $w(l)$. Clearly, $\sum_{l=1}^{c(w^n)} c_l(u^n|w^n) = c(u^n, w^n)$. In the above example, $w(1) = 0$, $w(2) = 1$, $w(3) = 01$, $c_1(u^6|w^6) = c_2(u^6|w^6) = 1$, and $c_3(u^6|w^6) = 2$. Now, the conditional LZ complexity of u^n given w^n is defined as

$$
\rho_{LZ}(u^n|w^n) \triangleq \frac{1}{n} \sum_{l=1}^{c(w^n)} c_l(u^n|w^n) \log c_l(u^n|w^n). \tag{26}
$$

We are now ready to present the main result of this section.

Theorem 3. *Consider the problem setting defined in Section 2 along with the above–mentioned modifications to incorporate side information. If there exists a stochastic encoder with q_e states and a decoder with q_d states that together satisfy the reliability constraint (9) and the security constraint (25), then its bandwidth expansion factor λ must be lower bounded as follows.*

$$
\lambda \geq \frac{\rho_{LZ}(u^n|w^n) - \Delta(\epsilon_r) - \epsilon_s - \eta_n(q_e \cdot q_d, k)}{C_s}, \tag{27}
$$

where

$$\eta_n(q_e \cdot q_d, k) = \min_{\{\ell \text{ divides } n/k\}} \left[\frac{\log(q_d q_e) + 1}{k\ell} + \frac{\log(4A^2)}{(1 - \epsilon_n) \log n} + \frac{A^2 \log(4A^2)}{n} \right], \tag{28}$$

with $\epsilon_n \to 0$ as $n \to \infty$ and $A = [(\alpha\omega)^{k\ell+1} - 1]/[\alpha\omega - 1]$, ω being the size of $\mathcal{W}$.

Note that the lower bound of Theorem 3 does not depend on the noisy side information at the wiretapper or on the channel $Q_{\dot{W}|W}$ that generates it from w^n. It depends only on u^n and w^n in terms of the data available in the system. Clearly, as it is a converse theorem, if it allows the side information to be available also at the encoder, then it definitely applies also to the case where the encoder does not have access to w^n. Interestingly, the encoder and the legitimate decoder act as if the wiretapper has the *clean* side information, w^n. While it is quite obvious that protection against availability of w^n at the wiretapper is sufficient for protection against availability of $\dot{W}^n$ (as $\dot{W}^n$ is a degraded version of w^n), it is not quite trivial that this should be also *necessary*, as the above converse theorem asserts. It is also interesting to note that here, the bound depends also on q_e, and not only q_d, as in Theorem 1. However, this dependence on q_e disappears in the special case where $\dot{W}^n = w^n$ with probability one.

We next discuss the achievability of the lower bound of Theorem 3. If the encoder has access to w^n, then the first step would be to apply the conditional LZ algorithm (see ([23], proof of Lemma 2) [24]), thus compressing u^n to about $n\rho_{LZ}(u^n|w^n)$ bits. The second step would be good channel coding for the wiretap channel, using the same methods as described in the previous section. If, however, the encoder does not have access to w^n, the channel coding part is still as before, but the situation with the source coding part is somewhat more involved since neither the encoder nor the decoder can calculate the target bit rate, $\rho_{LZ}(u^n|w^n)$, as neither party has access to both u^n and w^n. However, this source coding rate can essentially be achieved, provided that there is a low-rate noiseless feedback channel from the legitimate decoder to the encoder. The following scheme is in the spirit of the one proposed by Draper [25], but with a few modifications.

The encoder implements random binning for all source sequences in $\mathcal{U}^n$, that is, for each member of $\mathcal{U}^n$ an index is drawn independently, under the uniform distribution over $\{0, 1, 2, \ldots, \alpha^n - 1\}$, which is represented by its binary expansion, $b(u^n)$, of length $n \log \alpha$ bits. We select a large positive integer r, but keep $r \ll n$ (say, $r = \sqrt{n}$ or $r = \log^2 n$). The encoder transmits the bits of $b(u^n)$ incrementally, r bits at a time, until it receives from the decoder ACK. Each chunk of r bits is fed into a good channel code for the wiretap channel, at a rate slightly less than C_s. At the decoder side, this channel code is decoded (correctly, with high probability, for large r). Then, for each i ($i = 1, 2, \ldots$), after having decoded the i-th chunk of r bits of $b(u^n)$, the decoder creates the list $\mathcal{A}_i(u^n) = \{\dot{u}^n : [b(\dot{u}^n)]^{ir} = [b(u^n)]^{ir}\}$, where $[b(\dot{u}^n)]^l$ denotes the string formed by the first l bits of $b(\dot{u}^n)$. For each $\dot{u}^n \in \mathcal{A}_i(u^n)$, the decoder calculates $\rho_{LZ}(\dot{u}^n|w^n)$. We fix an arbitrarily small $\delta > 0$, which controls the trade-off between error probability and compression rate. If $n\rho_{LZ}(\dot{u}^n|w^n) \leq i \cdot r - n\delta$ for some $\dot{u}^n \in \mathcal{A}_i(u^n)$, the decoder sends ACK on the feedback channel and outputs the reconstruction, $\dot{u}^n$, with the smallest $\rho_{LZ}(\dot{u}^n|w^n)$ among all members of $\mathcal{A}_i(u^n)$. If no member of $\mathcal{A}_i(u^n)$ satisfies $n\rho_{LZ}(\dot{u}^n|w^n) \leq i \cdot r - n\delta$, the receiver waits for the next chunk of r compressed bits, and it does not send ACK. The probability of source-coding error after the i-th chunk is upper bounded by

$$
\begin{aligned}
P_e(i) &\overset{(a)}{\leq} |\{\dot{u}^n \neq u^n : n\rho_{LZ}(\dot{u}^n|w^n) \leq i \cdot r - n\delta\}| \cdot 2^{-i \cdot r} \\
&\overset{(b)}{\leq} \exp_2\left\{ i \cdot r - n\delta + O\left(\frac{n \log(\log n)}{\log n}\right)\right\} \cdot 2^{-i \cdot r} \\
&= \exp_2\left\{-n\delta + O\left(\frac{n \log(\log n)}{\log n}\right)\right\} \\
&\to 0 \quad \text{as } n \to \infty,
\end{aligned}
\tag{29}
$$

where in (a), the factor $2^{-i\cdot r}$ is the probability that $[b(\dot{u}^n)]^{ir} = [b(u^n)]^{ir}$ for each member of the set $\{\dot{u}^n \neq u^n : \ n\rho_{\text{LZ}}(\dot{u}^n|w^n) \leq i \cdot r - n\delta\}$ and (b) is based on ([23], Equation (A.13)). Clearly, it is guaranteed that an ACK is received at the encoder (and hence the transmission stops), no later than after the transmission of chunk no. i^*, where i^* is the smallest integer i such that $i \cdot r \geq n\rho_{\text{LZ}}(u^n|w^n) + n\delta$, namely, $i^* = \lceil [n\rho_{\text{LZ}}(u^n|w^n) + n\delta]/r \rceil$, which is the stage at which at least the correct source sequence begins to satisfy the condition $n\rho_{\text{LZ}}(u^n|w^n) \leq i \cdot r - n\delta$. Therefore, the compression ratio is no worse than $i^* \cdot r/n = \lceil n[\rho_{\text{LZ}}(u^n|w^n) + \delta]/r \rceil \cdot r/n \leq \rho_{\text{LZ}}(u^n|w^n) + \delta + r/n$. The overall probability of source-coding error is then upper bounded by

$$P_{\text{e}} = \Pr \bigcup_{i=1}^{i^*} \{\text{error at state } i\} \leq \sum_{i=1}^{i^*} P_{\text{e}}(i) \leq \left(\frac{n \log \alpha}{r} + 1 \right) \cdot \exp_2 \left\{ -n\delta + O\left(\frac{n \log(\log n)}{\log n} \right) \right\}, \tag{30}$$

which still tends to zero as $n \to \infty$. As for channel-coding errors, the probability that at least one chunk is decoded incorrectly is upper bounded by $(\frac{n \log \alpha}{r} + 1) \cdot e^{-rE}$, where E is an achievable error exponent of channel coding at the given rate. Thus, if r grows at any rate faster than logarithmic, but sub-linearly in n, then the overall channel-coding error probability tends to zero and, at the same time, the compression redundancy, r/n, tends to zero too.

To show that the security constraint (25) is satisfied too, consider an arbitrary assignment μ of random vectors (U^n, W^n), and let us denote by B the string of $I(X^N; Z^N) - N\epsilon$ bits of local randomness in Wyner's code [1]. Then,

$$
\begin{aligned}
I(X^N; Z^N) &= H(Z^N) - H(Z^N|X^N) \\
&\overset{(a)}{\geq} H(Z^N) - H(Z^N|U^n, B) \\
&\overset{(b)}{\geq} H(Z^N|\dot{W}^n) - H(Z^N|U^n, B) \\
&\overset{(c)}{=} H(Z^N|\dot{W}^n) - H(Z^N|U^n, B, \dot{W}^n) \\
&= I(U^n, B; Z^N|\dot{W}^n) \\
&= H(U^n, B|\dot{W}^n) - H(U^n, B|Z^N, \dot{W}^n) \\
&= H(U^n|\dot{W}^n) + H(B|U^n, \dot{W}^n) - H(U^n|Z^N, \dot{W}^n) - H(B|Z^N, \dot{W}^n, U^n) \\
&\overset{(d)}{=} H(U^n|\dot{W}^n) + H(B) - H(U^n|Z^N, \dot{W}^n) - H(B|Z^N, \dot{W}^n, U^n) \\
&\overset{(e)}{\geq} H(U^n|\dot{W}^n) + H(B) - H(U^n|Z^N, \dot{W}^n) - H(B|Z^N, U^n) \\
&\overset{(f)}{\geq} H(U^n|\dot{W}^n) + [I(X^N; Z^N) - N\epsilon] - H(U^n|Z^N, \dot{W}^n) - n\delta_n \\
&= I(X^N; Z^N) + I_\mu(U^n; Z^N|\dot{W}^n) - n(\lambda\epsilon + \delta_m),
\end{aligned}
\tag{31}
$$

where (a) is due to $(U^n, B) \to X^N \to Z^N$ being a Markov chain, (b) is due to conditioning reducing entropy, (c) is due to $\dot{W}^n \to (U^n, B) \to Z^N$ being a Markov chain, (d) is due to B being independent of $(U^n, \dot{W}^n)$, (e) is due to conditioning reducing entropy, and (f) is due to, in Wyner coding, B being able to be reliably decoded given that (Z^N, U^n) (δ_n is understood to be small, and recall that W^n is not needed in the channel decoding phase, but only in the Slepian–Wold decoding phase), and that the length of B is chosen to be $I(X^N; Z^N) - N\epsilon$. Comparing the right-most side to the left-most side, we readily obtain

$$I_\mu(U^n; Z^N|\dot{W}^n) \leq n(\lambda\epsilon + \delta_n), \tag{32}$$

which can be made arbitrarily small.

5. Proofs

We begin this section by establishing more notation conventions to be used throughout all proofs.

Let $n \gg k$ be a positive integer and let ℓ be such that $K \overset{\triangle}{=} \ell \cdot k$ divides n. Consider the partition of u^n into n/K non-overlapping blocks of length K,

$$
\begin{aligned}
&(\tilde{u}_0, \tilde{u}_1, \ldots, \tilde{u}_{\ell-1}), (\tilde{u}_\ell, \tilde{u}_{\ell+1}, \ldots, \tilde{u}_{2\ell-1}), \ldots, (\tilde{u}_{n/k-\ell}, \tilde{u}_{n/k-\ell+1}, \tilde{u}_{n/k-1}) \\
&= (u_1^K, u_{K+1}^{2K}, \ldots, u_{n-K+1}^n)
\end{aligned}
\tag{33}
$$

and apply the same partition to v^n. The corresponding channel input and output sequences are of length $N = n\lambda$. Let $M = \ell \cdot m = K\lambda$ and consider the parallel partition of the channels input and output sequences according to

$$
\begin{aligned}
&(\tilde{x}_0, \tilde{x}_1, \ldots, \tilde{x}_{\ell-1}), (\tilde{x}_\ell, \tilde{x}_{\ell+1}, \ldots, \tilde{x}_{2\ell-1}), \ldots, (\tilde{x}_{N/m-\ell}, \tilde{x}_{N/m-\ell+1}, \ldots, \tilde{x}_{N/m-1}) \\
&(\tilde{y}_0, \tilde{y}_1, \ldots, \tilde{y}_{\ell-1}), (\tilde{y}_\ell, \tilde{y}_{\ell+1}, \ldots, \tilde{y}_{2\ell-1}), \ldots, (\tilde{y}_{N/m-\ell}, \tilde{y}_{N/m-\ell+1}, \ldots, \tilde{y}_{N/m-1}) \\
&(\tilde{z}_0, \tilde{z}_1, \ldots, \tilde{z}_{\ell-1}), (\tilde{z}_\ell, \tilde{z}_{\ell+1}, \ldots, \tilde{z}_{2\ell-1}), \ldots, (\tilde{z}_{N/m-\ell}, \tilde{z}_{N/m-\ell+1}, \ldots, \tilde{z}_{N/m-1}).
\end{aligned}
\tag{34}
$$

For the sake of brevity, we henceforth denote $(\tilde{u}_{i\ell}, \ldots, \tilde{u}_{(i+1)\ell-1})$ by $\tilde{u}_{i\ell}^{(i+1)\ell-1}$ and use the same notation rule for all other sequences. Next, define the joint empirical distribution

$$
\begin{aligned}
&P_{\hat{U}^K \hat{X}^M \hat{Y}^M \hat{Z}^M \hat{S}^{\mathrm{e}} \hat{S}^{\mathrm{d}}}(u^K, x^M, y^M, z^M, s^{\mathrm{e}}, s^{\mathrm{d}}) = \\
&\frac{K}{n} \sum_{i=0}^{n/K-1} \delta\{\tilde{u}_{i\ell}^{(i+1)\ell-1} = u^K, \tilde{x}_{i\ell}^{(i+1)\ell-1} = x^M, \tilde{y}_{i\ell}^{(i+1)\ell-1} = y^M, \\
&\tilde{z}_{i\ell}^{(i+1)\ell-1} = z^M, s_{i\ell+1}^{\mathrm{e}} = s^{\mathrm{e}}, s_{i\ell+1}^{\mathrm{d}} = s^{\mathrm{d}}\},
\end{aligned}
\tag{35}
$$

and

$$
P_{\hat{U}^K X^M Y^M Z^M \hat{S}^{\mathrm{e}} S^{\mathrm{d}}}(u^K, x^M, y^M, z^M, s^{\mathrm{e}}, s^{\mathrm{d}}) = \mathbb{E}\left\{ P_{\hat{U}^K \hat{X}^M \hat{Y}^M \hat{Z}^M \hat{S}^{\mathrm{e}} \hat{S}^{\mathrm{d}}}(u^K, x^M, y^M, z^M, s^{\mathrm{e}}, s^{\mathrm{d}}) \right\},
\tag{36}
$$

where the expectation is w.r.t. both the randomness of the encoder and the randomness of both channels. Note that

$$
P_{\hat{U}^K X^M Y^M Z^M \hat{S}^{\mathrm{e}}}(u^K, x^M, y^M, z^M, s^{\mathrm{e}}) = P_{\hat{U}^K \hat{S}^{\mathrm{e}}}(u^K, s^{\mathrm{e}}) P(x^M | u^K, s^{\mathrm{e}}) Q_{\mathrm{M}}(y^M | x^M) Q_{\mathrm{W}}(z^M | y^M).
\tag{37}
$$

where

$$
P(x^M | u^K, s^{\mathrm{e}}) = \prod_{j=0}^{\ell-1} P(\tilde{x}_j | \tilde{u}_j, s_j^{\mathrm{e}}), \quad s_0^{\mathrm{e}} = s^{\mathrm{e}}
\tag{38}
$$

$$
Q_{\mathrm{M}}(y^M | x^M) = \prod_{j=0}^{M-1} Q_{\mathrm{M}}(y_i | x_i)
\tag{39}
$$

$$
Q_{\mathrm{W}}(z^M | y^M) = \prod_{j=0}^{M-1} Q_{\mathrm{M}}(z_i | y_i).
\tag{40}
$$

Note also that the bit error probability (in the absence of side information) under this distribution is

$$
\frac{1}{K}\mathbb{E}\{d_{\mathrm{H}}(\hat{U}^K, f(Y^M, S^{\mathrm{d}}))\}
$$

$$
= \frac{1}{K}\sum_{u^K, y^M, s^{\mathrm{e}}, s^{\mathrm{d}}} P_{\hat{U}^K Y^M S^{\mathrm{e}} S^{\mathrm{d}}}(u^K, y^M, s^{\mathrm{e}}, s^{\mathrm{d}}) d_{\mathrm{H}}(u^K, f(y^M, s^{\mathrm{d}}))
$$

$$
= \frac{1}{K}\sum_{u^K, y^M, s^{\mathrm{d}}} \frac{K}{n}\sum_{i=0}^{n/K-1} \mathbb{E}\left[\delta\{\tilde{u}_{i\ell}^{(i+1)\ell-1} = u^K, s_{i\ell+1}^{\mathrm{e}} = s^{\mathrm{e}}, \tilde{y}_{i\ell}^{(i+1)\ell-1} = y^M, s_{i\ell+1}^{\mathrm{d}} = s^{\mathrm{d}}\right] \times \quad (41)
$$

$$
d_{\mathrm{H}}(u^K, f(y^M, s^{\mathrm{d}}))
$$

$$
= \frac{1}{n}\sum_{i=0}^{n/K-1}\sum_{y^M, s^{\mathrm{d}}} F(y^M, s^{\mathrm{d}}|u_{iK+1}^{iK+K}, s_{i\ell+1}^{\mathrm{e}}) d_{\mathrm{H}}(u_{iK+1}^{iK+K}, f(y^M, s^{\mathrm{d}}))
$$

$$
= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\{d_{\mathrm{H}}(u_i, V_i)\},
$$

where $f(Y^M, S^{\mathrm{d}})$ is induced by ℓ successive applications of the decoder output function with inputs $Y^m, Y_{m+1}^{2m}, \ldots, Y_{M-m+1}^M$ and the initial state S^{d}, and where

$$
F(y^M, s^{\mathrm{d}}|u^K, s^{\mathrm{e}}) = \sum_{x^M} P(x^M|u^K, s^{\mathrm{e}}) Q_{\mathrm{M}}(y^M|x^M) P_{S^{\mathrm{d}}|Y^M}(s^{\mathrm{d}}|y^M). \quad (42)
$$

5.1. Proof of Theorem 1

Beginning with the reliability constraint, we have

$$
\begin{aligned}
I(\hat{U}^K; Y^M, S^{\mathrm{d}}) &= H(\hat{U}^K) - H(\hat{U}^K|Y^M, S^{\mathrm{d}}) \\
&= H(\hat{U}^K) - H(\hat{U}^K|Y^M) + I(S^{\mathrm{d}}; \hat{U}^K|Y^M) \\
&\leq I(\hat{U}^K; Y^M) + H(S^{\mathrm{d}}|Y^M) \\
&\leq I(X^M; Y^M) + \log q_{\mathrm{d}}.
\end{aligned} \quad (43)
$$

On the other hand,

$$
\begin{aligned}
I(\hat{U}^K; Y^M, S^{\mathrm{d}}) &= H(\hat{U}^K) - H(\hat{U}^K|Y^M, S^{\mathrm{d}}) \\
&\geq H(\hat{U}^K) - K\Delta(\epsilon_{\mathrm{r}}),
\end{aligned} \quad (44)
$$

and so,

$$
\begin{aligned}
I(X^M; Y^M) &\geq H(\hat{U}^K) - K\Delta(\epsilon_{\mathrm{r}}) - \log q_{\mathrm{d}} \\
&\overset{\triangle}{=} K \cdot R(u^n, q_{\mathrm{d}}, \epsilon_{\mathrm{r}}) \\
&= M \cdot \frac{R(u^n, q_{\mathrm{d}}, \epsilon_{\mathrm{r}})}{\lambda}.
\end{aligned} \quad (45)
$$

Following [1], we define the function

$$
\Gamma[R] = \max_{\{P_X:\, I(X;Y) \geq R\}} I(X;Y|Z) = \max_{\{P_X:\, I(X;Y) \geq R\}} [I(X;Y) - I(X;Z)], \quad (46)
$$

which is monotonically non–increasing and concave ([1], Lemma 1). Regarding the security constraint,

$$
\begin{aligned}
H(\hat{U}^K) - K\epsilon_{\rm s} \;&\overset{(a)}{\leq}\; H(\hat{U}^K) - \max_{\mu} I_{\mu}(U^K; Z^M) \\
&\leq\; H(\hat{U}^K) - I(\hat{U}^K; Z^M) \\
&=\; H(\hat{U}^K|Z^M) - H(\hat{U}^K|Y^M, Z^M, S^{\rm d}) + H(\hat{U}^K|Y^M, Z^M, S^{\rm d}) \\
&=\; H(\hat{U}^K|Z^M) - H(\hat{U}^K|Y^M, Z^M) + I(S^{\rm d}; \hat{U}^K|Y^M, Z^M) + H(\hat{U}^K|Y^M, Z^M, S^{\rm d}) \\
&\overset{(b)}{\leq}\; I(\hat{U}^K; Y^M|Z^M) + \log q_{\rm d} + K\Delta(\epsilon_{\rm r}) \\
&\overset{(c)}{\leq}\; I(X^M; Y^M|Z^M) + \log q_{\rm d} + K\Delta(\epsilon_{\rm r}) \\
&\overset{(d)}{\leq}\; \sum_{i=1}^{M} I(X_i; Y_i|Z_i, Y^{i-1}) + \log q_{\rm d} + K\Delta(\epsilon_{\rm r}) \\
&=\; \sum_{i=1}^{M} \sum_{y^{i-1}} P_{Y^{i-1}}(y^{i-1}) I(X_i; Y_i|Z_i, Y^{i-1} = y^{i-1}) + \log q_{\rm d} + K\Delta(\epsilon_{\rm r}) \\
&\overset{(e)}{\leq}\; M \cdot \frac{1}{M} \sum_{i=1}^{M} \sum_{y^{i-1}} P_{Y^{i-1}}(y^{i-1}) \Gamma[I(X_i; Y_i|Y^{i-1} = y^{i-1})] + \log q_{\rm d} + K\Delta(\epsilon_{\rm r}) \\
&\overset{(f)}{\leq}\; M \cdot \Gamma\left[\frac{1}{M} \sum_{i=1}^{M} \sum_{y^{i-1}} P_{Y^{i-1}}(y^{i-1}) I(X_i; Y_i|Y^{i-1} = y^{i-1})\right] + \log q_{\rm d} + K\Delta(\epsilon_{\rm r}) \\
&=\; M \cdot \Gamma\left[\frac{1}{M} \sum_{i=1}^{M} I(X_i; Y_i|Y^{i-1})\right] + \log q_{\rm d} + K\Delta(\epsilon_{\rm r}) \\
&=\; M \cdot \Gamma\left[\frac{1}{M} \sum_{i=1}^{M} \left\{ H(Y_i|Y^{i-1}) - H(Y_i|X_i, Y^{i-1}) \right\}\right] + \log q_{\rm d} + K\Delta(\epsilon_{\rm r}) \\
&=\; M \cdot \Gamma\left[\frac{1}{M} \sum_{i=1}^{M} \left\{ H(Y_i|Y^{i-1}) - H(Y_i|X_i) \right\}\right] + \log q_{\rm d} + K\Delta(\epsilon_{\rm r}) \\
&=\; M \cdot \Gamma\left[\frac{1}{M} \left\{ H(Y^M) - H(Y^M|X^M) \right\}\right] + \log q_{\rm d} + K\Delta(\epsilon_{\rm r}) \\
&=\; M \cdot \Gamma\left[\frac{I(X^M; Y^M)}{M}\right] + \log q_{\rm d} + K\Delta(\epsilon_{\rm r}) \\
&\overset{(g)}{\leq}\; M \cdot \Gamma\left[\frac{R(u^n, q_{\rm d}, \epsilon_{\rm r})}{\lambda}\right] + \log q_{\rm d} + K\Delta(\epsilon_{\rm r}) \\
&\leq\; M \cdot \Gamma\left[\frac{R(u^n, q_{\rm d}, \epsilon_{\rm r}) - \epsilon_{\rm s}}{\lambda}\right] + \log q_{\rm d} + K\Delta(\epsilon_{\rm r}),
\end{aligned}
\tag{47}
$$

where $P_{Y^{i-1}}(y^{i-1}) = \sum_{y_i^M} P_{Y^M}(y^M)$, (a) is due to the security constraint, (b) follows from Fano's inequality and the fact that $I(S^{\rm d}; \hat{U}^K|Y^M, Z^M) \leq H(S^{\rm d}) \leq \log q_{\rm d}$, (c) is due to the data processing inequality and the fact that $\hat{U}^K \to X^M \to Y^M$ is a Markov chain given Z^M, (d) is as in ([1], Equation (37)), (e) is by the definition of Wyner's function $\Gamma(\cdot)$, (f) is by the concavity of this function, and (g) is by (45) and the decreasing monotonicity of the function $\Gamma(\cdot)$. Thus,

$$
\frac{R(u^n, q_{\rm d}, \epsilon_{\rm r}) - \epsilon_{\rm s}}{\lambda} \leq \Gamma\left[\frac{R(u^n, q_{\rm d}, \epsilon_{\rm r}) - \epsilon_{\rm s}}{\lambda}\right]
\tag{48}
$$

or

$$
\frac{R(u^n, q_{\rm d}, \epsilon_{\rm r}) - \epsilon_{\rm s}}{\lambda} \leq C_{\rm s}
\tag{49}
$$

which is

$$
R(u^n, q_{\rm d}, \epsilon_{\rm r}) \leq \lambda C_{\rm s} + \epsilon_{\rm s}
\tag{50}
$$

or, equivalently,

$$\frac{H(\hat{U}^K)}{K} \le \lambda C_{\mathrm{s}} + \epsilon_{\mathrm{s}} + \Delta(\epsilon_{\mathrm{r}}) + \frac{\log q_{\mathrm{d}}}{K}. \tag{51}$$

Finally, we apply the inequality ([20], Equation (18))

$$\frac{H(\hat{U}^K)}{K} \ge \rho_{\mathrm{LZ}}(u^n) - \frac{2K(\log \alpha + 1)^2}{(1 - \epsilon_n)\log n} - \frac{2K\alpha^{2K}\log \alpha}{n} - \frac{1}{K}, \tag{52}$$

to obtain

$$\rho_{\mathrm{LZ}}(u^n) \le \lambda C_{\mathrm{s}} + \epsilon_{\mathrm{s}} + \Delta(\epsilon_{\mathrm{r}}) + \zeta_n(q_{\mathrm{d}}, k), \tag{53}$$

which completes the proof of Theorem 1.

5.2. Proof of Theorem 2

Consider the following extension of the joint distribution to include a random variable that represents $\{b_i\}$, as follows:

$$P_{\hat{U}^K B^J X^M Y^M Z^M \hat{S}^{\mathrm{e}} S^{\mathrm{d}}}(u^K, b^J, x^M, y^M, z^M, s^{\mathrm{e}}, s^{\mathrm{d}}) =$$

$$\frac{K}{n} \sum_{i=0}^{n/K-1} \mathbb{E}\left[\delta\{\tilde{u}_{i\ell}^{(i+1)\ell-1} = u^K, \tilde{b}_{i\ell}^{(i+1)\ell-1} = b^J, \tilde{x}_{i\ell}^{(i+1)\ell-1} = x^M, \tilde{y}_{i\ell}^{(i+1)\ell-1} = y^M, \tag{54}\right.$$

$$\left. \tilde{z}_{i\ell}^{(i+1)\ell-1} = z^M, s_{i\ell+1}^{\mathrm{e}} = s^{\mathrm{e}}, s_{i\ell+1}^{\mathrm{d}} = s^{\mathrm{d}}\}\right],$$

where $J = j\ell$ and $\tilde{b}_{i\ell}^{(i+1)\ell-1} = (\tilde{b}_{i\ell}, \tilde{b}_{i\ell+1}, \ldots, \tilde{b}_{(i+1)\ell-1})$. Next, consider the following chain of inequalities

$$\begin{aligned}
K\epsilon_{\mathrm{s}} &\ge \max_{\mu} I_{\mu}(U^K; Z^M) \\
&\ge I(\hat{U}^K; Z^M) \\
&= I(\hat{U}^K, B^J, S^{\mathrm{e}}; Z^M) - I(B^J, S^{\mathrm{e}}; Z^M | \hat{U}^K) \\
&\overset{\text{(a)}}{=} I(X^M; Z^M) - I(B^J, S^{\mathrm{e}}; Z^M | \hat{U}^K) \\
&\ge I(X^M; Z^M) - H(B^J, S^{\mathrm{e}} | \hat{U}^K) \\
&\ge I(X^M; Z^M) - H(B^J, S^{\mathrm{e}}) \\
&\ge I(X^M; Z^M) - H(B^J) - H(S^{\mathrm{e}}) \\
&\ge I(X^M; Z^M) - J - \log q_{\mathrm{e}},
\end{aligned} \tag{55}$$

where (a) is due to the fact that, on the one hand, X^M is a deterministic function of $(\hat{U}^K, B^J, S^{\mathrm{e}})$, which implies that $I(\hat{U}^K, B^J, S^{\mathrm{e}}; Z^M) \ge I(X^M; Z^M)$, but on the other hand, $(\hat{U}^K, B^J, S^{\mathrm{e}}) \to X^M \to Z^M$ is a Markov chain and so, $I(\hat{U}^K, B^J, S^{\mathrm{e}}; Z^M) \le I(X^M; Z^M)$, hence the equality. Thus,

$$J \ge I(X^M; Z^M) - K\epsilon_{\mathrm{s}} - \log q_{\mathrm{e}}, \tag{56}$$

or

$$j \ge \frac{I(X^M; Z^M)}{\ell} - k\epsilon_{\mathrm{s}} - \frac{\log q_{\mathrm{e}}}{\ell} \ge \frac{mI(X^M; Z^M)}{M} - k\epsilon_{\mathrm{s}} - \frac{\log q_{\mathrm{e}}}{\ell}. \tag{57}$$

The meaning of this result is the following: once one finds a communication system that complies with both the security constraint and the reliability constraint, then the amount of local randomization is lower bounded in terms of the induced mutual information, $I(X^M; Z^M)$, as above. By the hypothesis of Theorem 2, the secrecy capacity is saturated, and hence P_{X^M} must coincide with the product distribution, $[P_{X^*}]^M$, yielding $I(X^M; Z^M)/M = I(X^*; Z^*)$. Thus,

$$j \ge mI(X^*; Z^*) - k\epsilon_{\mathrm{s}} - \frac{\log q_{\mathrm{e}}}{\ell}. \tag{58}$$

This completes the proof of Theorem 2.

5.3. Outline of the Proof of Theorem 3

The proof follows essentially the same steps as those of the proof of Theorem 1, except that everything should be conditioned on the side information, but there are also some small twists. We, therefore, only provide a proof outline and highlight the differences.

The auxiliary joint distribution is now extended to read

$$
\begin{aligned}
& P_{\hat{U}^K \hat{W}^K \dot{W}^K X^N Y^N Z^N \hat{S}^e S^d}(u^K, w^K, \dot{w}^K, x^N, y^N, z^N, s^e, s^d) = \\
& \frac{K}{m} \sum_{i=0}^{m/K-1} \mathbb{E}\Big[\delta\{\tilde{u}_{i\ell}^{(i+1)\ell-1} = u^K, \tilde{w}_{i\ell}^{(i+1)\ell-1} = w^K, \tilde{\dot{w}}_{i\ell}^{(i+1)\ell-1} = \dot{w}^K, \tilde{x}_{i\ell}^{(i+1)\ell-1} = x^M, \quad (59) \\
& \tilde{y}_{i\ell}^{(i+1)\ell-1} = y^M, \tilde{z}_{i\ell}^{(i+1)\ell-1} = z^M, s_{i\ell+1}^e = s^e, s_{i\ell+1}^d = s^d\}\Big].
\end{aligned}
$$

Note that

$$
\begin{aligned}
& P_{\hat{U}^K \hat{W}^K \dot{W}^K Z^M \hat{S}^e}(u^k, w^k, \dot{w}^k, z^M, s^e) \\
& = \frac{K}{n} \sum_{i=0}^{n/K-1} \mathbb{E}\Big[\delta\{\tilde{u}_{i\ell}^{(i+1)\ell-1} = u^K, \tilde{w}_{i\ell}^{(i+1)\ell-1} = w^K, \tilde{\dot{w}}_{i\ell}^{(i+1)\ell-1} = \dot{w}^K, \tilde{z}_{i\ell}^{(i+1)\ell-1} = z^M, s_{i\ell+1}^e = s^e\}\Big] \qquad (60) \\
& = \frac{K}{n} \sum_{i=0}^{n/K-1} \delta\{\tilde{u}_{i\ell}^{(i+1)\ell-1} = u^K, \tilde{w}_{i\ell}^{(i+1)\ell-1} = w^K, s_{i\ell+1}^e = s^e\} \cdot Q_{\dot{W}^K|W^K}(\dot{w}^K|w^K) \cdot G(z^M|u^K, s^e) \\
& = P_{\hat{U}^K \hat{W}^K \hat{S}^e}(u^K, w^K, s^e) \cdot Q_{\dot{W}^K|W^K}(\dot{w}^K|w^K) \cdot G(z^M|u^K, s^e),
\end{aligned}
$$

where

$$
G(z^M|u^K, s^e) = \sum_{x^M} P(x^M|u^K, s^e) Q_{\text{MW}}(z^M|x^M). \qquad (61)
$$

It follows that $\dot{W}^K \to \hat{W}^K \to (\hat{U}^K, \hat{S}^e) \to Z^M$ is a Markov chain under $P_{\hat{U}^K \hat{W}^K \dot{W}^K Z^M \hat{S}^e}$. In other words, the legitimate decoder has side information of better quality than that of the wiretapper. First, observe that

$$
\begin{aligned}
I_\mu(U^n; Z^N|W^n) & = H_\mu(Z^N|W^n) - H_\mu(Z^N|W^n, U^n) \\
& \leq H_\mu(Z^N|W^n) - H_\mu(Z^N|W^n, U^n, S^e) \\
& = H_\mu(Z^N|W^n) - H_\mu(Z^N|U^n, S^e) \\
& \leq H_\mu(Z^N|\dot{W}^n) - H_\mu(Z^N|U^n, S^e) \qquad (62) \\
& = H_\mu(Z^N|\dot{W}^n) - H_\mu(Z^N|\dot{W}^n, U^n, S^e) \\
& \leq H_\mu(Z^N|\dot{W}^n) - H_\mu(Z^N|\dot{W}^n, U^n) + \log q_e \\
& = I_\mu(U^n; Z^N|\dot{W}^n) + \log q_e.
\end{aligned}
$$

The reliability constraint is handled exactly as in the proof of Theorem 1, except that everything should be conditioned on $\hat{W}^K$. The result of this is

$$
\begin{aligned}
I(X^M; Y^M|\hat{W}^K) & \geq H(\hat{U}^K|\hat{W}^K) - K\Delta(\epsilon_r) - \log q_d \\
& \triangleq K \cdot R(u^n, w^n, q_d, \epsilon_r) \qquad (63) \\
& = M \cdot \frac{R(u^n, w^n, q_d, \epsilon_r)}{\lambda}.
\end{aligned}
$$

Regarding the security constraint, we begin with the following manipulation.

$$
\begin{aligned}
H(\hat{U}^K | Z^M, \dot{W}^K) \;=\;& H(\hat{U}^K | \dot{W}^K) - I(\hat{U}^K; Z^M | \dot{W}^K) \\
=\;& H(\hat{U}^K | \dot{W}^K) - H(\hat{U}^K | \hat{W}^K) + H(\hat{U}^K | \hat{W}^K) - I(\hat{U}^K; Z^M | \dot{W}^K) \\
\overset{(a)}{\leq}\;& H(\hat{U}^K | \dot{W}^K) - H(\hat{U}^K | \hat{W}^K) + H(\hat{U}^K | \hat{W}^K) - I(\hat{U}^K; Z^M | \hat{W}^K) + \log q_{\mathrm{e}} \\
=\;& H(\hat{U}^K | \dot{W}^K) - H(\hat{U}^K | \hat{W}^K) + H(\hat{U}^K | Z^M, \hat{W}^K) + \log q_{\mathrm{e}} \\
=\;& H(\hat{U}^K | \dot{W}^K) - H(\hat{U}^K | \hat{W}^K) + H(\hat{U}^K | Z^M, \hat{W}^K) - \\
& H(\hat{U}^K | Y^M, Z^M, S^{\mathrm{d}}, \hat{W}^K) + H(\hat{U}^K | Y^M, Z^M, S^{\mathrm{d}}, \hat{W}^K) + \log q_{\mathrm{e}} \\
=\;& H(\hat{U}^K | \dot{W}^K) - H(\hat{U}^K | \hat{W}^K) + H(\hat{U}^K | Z^M, \hat{W}^K) - H(\hat{U}^K | Y^M, Z^M, \hat{W}^K) + \\
& I(S^{\mathrm{d}}; \hat{U}^K | Y^M, Z^M, \hat{W}^K) + H(\hat{U}^K | Y^M, Z^M, S^{\mathrm{d}}, \hat{W}^K) + \log q_{\mathrm{e}} \\
\overset{(b)}{\leq}\;& H(\hat{U}^K | \dot{W}^K) - H(\hat{U}^K | \hat{W}^K) + I(\hat{U}^K; Y^M | Z^M, \hat{W}^K) + \log q_{\mathrm{d}} + \\
& K\Delta(\epsilon_{\mathrm{r}}) + \log q_{\mathrm{e}} \\
\leq\;& H(\hat{U}^K | \dot{W}^K) - H(\hat{U}^K | \hat{W}^K) + I(\hat{U}^K, \hat{S}^{\mathrm{e}}; Y^M | Z^M, \hat{W}^K) + \log q_{\mathrm{d}} + \\
& K\Delta(\epsilon_{\mathrm{r}}) + \log q_{\mathrm{e}} \\
\overset{(c)}{\leq}\;& H(\hat{U}^K | \dot{W}^K) - H(\hat{U}^K | \hat{W}^K) + I(X^M; Y^M | Z^M, \hat{W}^K) + \log(q_{\mathrm{e}} q_{\mathrm{d}}) + K\Delta(\epsilon_{\mathrm{r}}),
\end{aligned}
\tag{64}
$$

where in (a) we used Equation (62), in (b) we used Fano's inequality, and in (c) we used the data processing inequality as $(\hat{U}^K, \hat{S}^{\mathrm{e}}) \to X^M \to Y^M$ is a Markov chain (also conditioned on $(\hat{W}^K, Z^M)$). The next step is to further upper bound the term $I(X^M; Y^M | Z^M, \hat{W}^K)$. This is carried out very similarly as in the proof of Theorem 1, except that everything is conditioned also on $\hat{W}^K$. We then obtain

$$
\begin{aligned}
H(\hat{U}^K | Z^M, \dot{W}^K) \;\leq\;& H(\hat{U}^K | \dot{W}^K) - H(\hat{U}^K | \hat{W}^K) + M \cdot \Gamma\left[\frac{R(u^n, w^n, q_{\mathrm{d}}, \epsilon_{\mathrm{r}})}{\lambda} \right] + \\
& \log(q_{\mathrm{e}} q_{\mathrm{d}}) + K\Delta(\epsilon_{\mathrm{r}}),
\end{aligned}
\tag{65}
$$

or, equivalently,

$$
\begin{aligned}
& H(\hat{U}^K | \hat{W}^K) - M \cdot \Gamma\left[\frac{R(u^n, w^n, q_{\mathrm{d}}, \epsilon_{\mathrm{r}})}{\lambda} \right] \\
\leq\;& H(\hat{U}^K | \dot{W}^K) - H(\hat{U}^K | Z^M, \dot{W}^K) + \log(q_{\mathrm{e}} q_{\mathrm{d}}) + K\Delta(\epsilon_{\mathrm{r}}) \\
=\;& I(\hat{U}^K; Z^M | \dot{W}^K) + \log(q_{\mathrm{e}} q_{\mathrm{d}}) + K\Delta(\epsilon_{\mathrm{r}}) \\
\leq\;& K\epsilon_{\mathrm{s}} + \log(q_{\mathrm{e}} q_{\mathrm{d}}) + K\Delta(\epsilon_{\mathrm{r}}),
\end{aligned}
\tag{66}
$$

or

$$
\begin{aligned}
R(u^n, w^n, q_{\mathrm{e}} \cdot q_{\mathrm{d}}, \epsilon_{\mathrm{r}}) \;\leq\;& \lambda \cdot \Gamma\left[\frac{R(u^n, w^n, q_{\mathrm{d}}, \epsilon_{\mathrm{r}})}{\lambda} \right] \\
\leq\;& \lambda \cdot \Gamma\left[\frac{R(u^n, w^n, q_{\mathrm{e}} \cdot q_{\mathrm{d}}, \epsilon_{\mathrm{r}})}{\lambda} \right]
\end{aligned}
\tag{67}
$$

which is the same as

$$
R(u^n, w^n, q_{\mathrm{e}} \cdot q_{\mathrm{d}}, \epsilon_{\mathrm{r}}) \leq \lambda \cdot C_{\mathrm{s}}.
\tag{68}
$$

or

$$
\frac{H(\hat{U}^K | \hat{W}^K)}{K} \leq \lambda \cdot C_{\mathrm{s}} + \epsilon_{\mathrm{s}} + \Delta(\epsilon_{\mathrm{r}}) + \frac{\log(q_{\mathrm{e}} \cdot q_{\mathrm{d}})}{K}.
\tag{69}
$$

The proof is completed by combining the last inequality with the following inequality ([26] Equations (17)–(19), [27] Equations (55)–(57)):

$$
\frac{H(\hat{U}^K | \hat{W}^K)}{K} \geq \rho_{\mathrm{LZ}}(u^n | w^n) - \frac{\log(4A^2)}{(1 - \epsilon_n) \log n} - \frac{A^2 \log(4A^2)}{n} - \frac{1}{K},
\tag{70}
$$

where $A = [(\alpha\omega)^{K+1} - 1]/[\alpha\omega - 1]$, ω being the alphabet size of $\mathcal{W}$.

References

1. Wyner, A.D. The wire-tap channel. *Bell Syst. Tech. J.* **1975**, *54*, 1355–1387. [CrossRef]
2. Csiszár, I.; Körner, J. Broadcast channels with confidential messages. *IEEE Trans. Inform. Theory* **1978**, *24*, 339–348. [CrossRef]
3. Leung-Yan-Cheong, S.K.; Hellman, M.E. The Gaussian wire-tap channel. *IEEE Trans. Inform. Theory* **1978**, *24*, 451–456. [CrossRef]
4. Ozarow, L.H.; Wyner, A.D. Wire–tap channel II. In Proceedings of the Eurocrypt 84, Workshop on Advances in Cryptology: Theory and Applications of Cryptographic Techniques, Paris, France, 9–11 April 1985; pp. 33–51.
5. Yamamoto, H. Coding theorems for secret sharing communication systems with two noisy channels. *IEEE Trans. Inform. Theory* **1989**, *35*, 572–578. [CrossRef]
6. Yamamoto, H. Rate–distortion theory for the Shannon cipher system. *IEEE Trans. Inform. Theory* **1997**, *43*, 827–835. [CrossRef]
7. Merhav, N. Shannon's secrecy system with informed receivers an its application to systematic coding for wiretapped channels. *IEEE Trans. Inform. Theory* **2008**, *54*, 2723–2734. [CrossRef]
8. Tekin, E.; Yener, A. The Gaussian multiple access wire–Tap channel. *IEEE Trans. Inform. Theory* **2008**, *54*, 5747–5755. [CrossRef]
9. Mitrpant, C. Information Hiding—An Application of Wiretap Channels with Side Information. Ph.D. Thesis, der Universitaet Duisburg–Essen, Essen, Germany, 2003.
10. Mitrpant, C.; Vinck, A.J.H.; Luo, Y. An achievable region for the Gaussian wiretap channel with side information. *IEEE Trans. Inform. Theory* **2006**, *52*, 2181–2190. [CrossRef]
11. Ardestanizadeh, E.; Franceschetti, M.; Javidi, T.; Kim, Y.-H. Wiretap channel with secure rate–Limited feedback. *IEEE Trans. Inform. Theory* **2009**, *55*, 5353–5361. [CrossRef]
12. Hayashi, M. Upper bounds of eavesdropper's performances in finite-length code with the decoy method. *Phy. Rev. A* **2007**, *76*, 012329. Erratum in *Phys. Rev. A* 2009, *79*, 019901E. [CrossRef]
13. Hayashi, M.; Matsumoto, R. Secure multiplex coding with dependent and non-uniform multiple messages. *IEEE Trans. Inform. Theory* **2016**, *62*, 2355–2409. [CrossRef]
14. Bloch, M.; Barros, J. *Physical-Layer Security: From Information Theory to Security Engineering*; Cambridge University Press: New York, NY, USA, 2011.
15. Bellare, M.; Tessaro, S.; Vardy, A. Semantic security for the wiretap channel. In *Advances in Cryptology—CRYPTO 2012*; Safavi-Naini, R., Canetti, R., Eds.; CRYPTO 2012 Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7417. [CrossRef]
16. Goldfeld, Z.; Cuff, P.; Permuter, H.H. Semantic security capacity for wiretap channels of type II. In Proceedings of the 2016 IEEE International Symposium on Information Theory (ISIT 2016), Barcelona, Spain, 10–15 July 2016; pp. 2799–2803.
17. Ziv, J. Coding theorems for individual sequences. *IEEE Trans. Inform. Theory* **1978**, *24*, 405–412. [CrossRef]
18. Ziv, J. Distortion–rate theory for individual sequences. *IEEE Trans. Inform. Theory* **1980**, *26*, 137–143. [CrossRef]
19. Ziv, J. Fixed–rate encoding of individual sequences with side information. *IEEE Trans. Inform. Theory* **1984**, *30*, 348–352. [CrossRef]
20. Merhav, N. Perfectly secure encryption of individual sequences. *IEEE Trans. Inform. Theory* **2013**, *58*, 1302–1310. [CrossRef]
21. Merhav, N. On the data processing theorem in the semi-deterministic setting. *IEEE Trans. Inform. Theory* **2014**, *60*, 6032–6040. [CrossRef]
22. Ziv, J.; Lempel, A. Compression of individual sequences via variable–Rate coding. *IEEE Trans. Inform. Theory* **1978**, *24*, 530–536. [CrossRef]
23. Ziv, J. Universal decoding for finite-state channels. *IEEE Trans. Inform. Theory* **1985**, *31*, 453–460. [CrossRef]
24. Uyematsu, T.; Kuzuoka, S. Conditional Lempel–Ziv complexity and its application to source coding theorem with side information. *IEICE Trans. Fundam.* **2003**, *E86-A*, 2615–2617.
25. Draper, S. Universal incremental Slepian–Wolf Coding. In Proceedings of the 43rd Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, USA, 29 September–1 October 2004; pp. 1332–1341.
26. Merhav, N. Universal detection of messages via finite–State channels. *IEEE Trans. Inform. Theory* **2000**, *46*, 2242–2246. [CrossRef]
27. Merhav, N. Guessing individual sequences: Generating randomized guesses using finite—State machines. *IEEE Trans. Inform. Theory* **2020**, *66*, 2912–2920. [CrossRef]