1 Effective Secrecy: Reliability, Confusion, and Stealth

Jie Hou, Gerhard Kramer, and Matthieu Bloch

A security measure called effective secrecy is defined that includes strong secrecy and stealth communication. Effective secrecy ensures that a message cannot be deciphered and that the presence of meaningful communication is hidden. To measure stealth we use resolvability via informational divergence and relate this to binary hypothesis testing. Results are developed for wiretap channels.

1.1 Introduction

The wiretap channel is depicted in Fig. 1.1 and has a message M that should be decoded reliably at one receiver (Bob) while being kept secret from a second receiver (Eve). Wyner [1] derived the *secrecy capacity* when the channel $P_{YZ|X}$ is *physically degraded*, i.e., X–Y–Z forms a Markov chain. Csiszár and Körner [2] extended the results to broadcast channels with confidential messages. In both [1] and [2], secrecy is measured by a *normalized* mutual information between M and Eve's output string $Z^n = Z_1 Z_2 \dots Z_n$, i.e., the secrecy requirement is

$$\frac{1}{n} I(M;Z^n) \leq S, \tag{1.1}$$

where we interpret S as a *leakage* rate. An interesting case is to choose S positive and small, in which case the requirement (1.1) is referred to as *weak secrecy*. However, as $n \to \infty$ the eavesdropper can obtain nS bits of M, which grows with n.

Instead, the papers [3, 4] advocated using *strong secrecy* where secrecy is measured by the *unnormalized* mutual information $I(M;Z^n)$ and one requires

$$I(M;Z^n) \leq \xi \tag{1.2}$$

for any $\xi > 0$ and sufficiently large n. We remark that Wyner's random codes already ensured strong secrecy since S in (1.1) can scale inverse-exponentially with n. Also, for *finite* n, whether we use (1.1) with $S = \xi/n$ or (1.2) is obviously immaterial, i.e., the distinction between weak and strong secrecy is of asymptotic nature only.

In related work, Han and Verdú [5] studied *resolvability* based on *variational distance* that addresses the number of bits needed to mimic a marginal distribution of a prescribed joint distribution. Hayashi [6] and Bloch and Laneman [7] used resolvability to prove secrecy, and they extended results in [2] to continuous random variables and channels

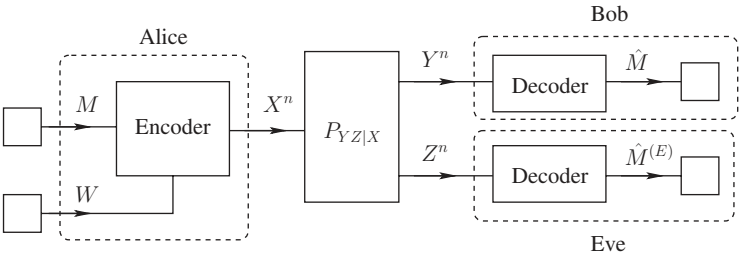

Figure 1.1 A wiretap channel with message M and random symbol W.

with memory. We also use the resolvability-based approach but replace variational distance by informational divergence (or Kullback–Leibler divergence).

The main contribution of this work is to define and justify the usefulness of a security measure that includes not only reliability and strong secrecy but also *stealth*. In particular, we measure secrecy by the informational divergence

$$D(P_{MZ^n} \| P_M Q_{Z^n}), \tag{1.3}$$

where P_{MZ^n} is the joint distribution of MZ^n, P_M is the distribution of M, P_{Z^n} is the distribution of Z^n, and Q_{Z^n} is the distribution that the eavesdropper expects to observe when the source is *not* communicating useful messages. We call this security measure *effective secrecy*. We show that classic random codes achieve effective secrecy by using a recently developed simplified proof [8] of resolvability based on informational divergence (see also [9, Lemma 11]).

It turns out that the effective secrecy measure (1.3) was considered a few months before our work [10] by Han, Endo, and Sasaki [11, 12]. Their motivation for using (1.3) was simply that it gives a secrecy measure that is stronger than strong secrecy. Our motivation was operational: the divergence (1.3) measures secrecy and stealth simultaneously. In particular, one can check that (see (1.7) below)

$$D(P_{MZ^n} \| P_M Q_{Z^n}) = \underbrace{I(M;Z^n)}_{\text{secrecy measure}} + \underbrace{D(P_{Z^n} \| Q_{Z^n})}_{\text{stealth measure}}, \tag{1.4}$$

where we measure secrecy and stealth by using $I(M;Z^n)$ and $D(P_{Z^n} \| Q_{Z^n})$, respectively. We justify the latter measure by using binary hypothesis testing in Section 1.4. Thus, by making $D(P_{MZ^n} \| P_M Q_{Z^n}) \to 0$ we not only keep the message secret from the eavesdropper but also hide the presence of meaningful communication. Of course, one can instead study secrecy and stealth separately rather than using (1.4); see Section 1.3.4 below. We combine these concepts mainly for convenience of the proofs.

The choice of default behavior Q_{Z^n} in (1.3) and (1.4) will depend on the application. For example, if the default behavior is to send a codeword, then $Q_{Z^n} = P_{Z^n}$ and one achieves stealth for free. On the other hand, if the default behavior is $Q_{Z^n} = Q_Z^n$, where Q_Z^n is a product distribution (see Section 1.2.2 below), then code design requires more care. We mostly focus on the case $Q_{Z^n} = Q_Z^n$.

This paper is organized as follows. In Section 1.2, we review terminology on low probability of detection and low probability of intercept communications. We further

describe notation and state the problem we study. In Section 1.3 we state and prove the main result. Section 1.4 relates this result to hypothesis testing, and Section 1.5 concludes the paper.

1.2 Preliminaries

1.2.1 Terminology

Many applications require hiding communication or information, and often the same concept is labeled with different words. We therefore begin by reviewing terminology in selected documents, and describe how we use the word *stealth*.

The United States Committee on National Security Systems Glossary [13] has the following definitions:

- Low probability of detection (LPD): Result of measures used to hide or disguise intentional electromagnetic transmissions.
- Low probability of intercept (LPI): Result of measures used to resist attempts by adversaries to analyze the parameters of a transmission to determine if it is a signal of interest.

The document [14, p. 6] has similar but slightly different terminology. There, LPI refers generically to communication methods whose primary purpose

> is to prevent an unauthorized listener from determining the presence or location of the transmitter, in order to decrease the possibility of both electronic attack (jamming) and physical attack.

The same document [14, p. 9] refers to [15] as describing

> four sequential operations that exploitation systems attempt to perform:
>
> 1. Cover the signal, that is, a receiver is tuned to some or all of the frequency intervals being occupied by the signal when the signal is actually being transmitted.
> 2. Detect the signal, that is, make a decision about whether the power in the intercept bandwidth is a signal plus noise and interference or just noise and interference.
> 3. Intercept the signal, that is, extract features of the signal to determine if it is a signal of interest or not.
> 4. Exploit the signal, that is, extract additional signal features as necessary and then demodulate the baseband signal to generate a stream of binary digits.

Secrecy deals with operation (4), i.e., the secrecy constraint prevents exploitation of the signal to generate a stream of (meaningful) binary digits. Our focus in this paper is on either operation (2) or (3), depending on how one interprets the above text. For example, suppose the default behavior Q_{Z^n} has Alice sending a signal whose power is (or more generally whose statistics are) sufficiently similar to interference. In this case, we are interested in operation (2). As a second example, suppose the default behavior Q_{Z^n} has Alice sending either a message-carrying signal or a default signal at irregular intervals with low probability. In this case, it does not matter if Eve detects that a signal was transmitted as long as she cannot determine if it is a message-carrying signal or not. We are thus concerned with operation (3).

We generically refer to scenarios in which message-carrying signals are shaped to resemble an innocent signal as *stealth* communication. The extreme scenario where signals are shaped to hide in noise has been referred to as covert communication [16], deniable communication [17], and undetectable communication [18]. This extreme type of stealth usually requires that X^n is a zero-power string so that no positive communication rate is possible. More precisely, the number of bits that can be communicated reliably over a noisy channel (often) scales as $\sqrt{n}$ [16, 17, 19, 20]. However, as we have seen, stealth communication rates are positive if random strings such as codewords are sent even if no information transmission occurs.

1.2.2 Notation

Random variables are written with upper case letters and their realizations with the corresponding lower case letters. Superscripts denote strings of variables/symbols, e.g., $X^n = X_1X_2\dots X_n$. Subscripts denote the position of a variable/symbol in a string. For instance, X_i denotes the ith variable in X^n. We use X_i^n to denote $X_i,\dots,X_n$, $1 \le i \le n$. A random variable X has probability distribution P_X and the support of P_X is denoted as $\mathrm{supp}(P_X)$. We write probabilities with subscripts $P_X(x)$ but we drop the subscripts if the arguments of the distribution are lower case versions of the random variables. For example, we write $P(x) = P_X(x)$. If the X_i, $i = 1,\dots,n$, are independent and identically distributed (i.i.d.) according to P_X, then we have $P(x^n) = \prod_{i=1}^n P_X(x_i)$ and we write $P_{X^n} = P_X^n$. We also use Q_X^n to refer to strings of i.i.d. random variables. Calligraphic letters denote sets. The size of a set $\mathcal{S}$ is denoted as $|\mathcal{S}|$ and the complement is denoted as $\mathcal{S}^c$. For X with alphabet $\mathcal{X}$, we denote $P_X(\mathcal{S}) = \sum_{x\in\mathcal{S}} P_X(x)$ for any $\mathcal{S} \subseteq \mathcal{X}$. We use $\mathcal{T}_\epsilon^n(P_X)$ to denote the set of letter-typical strings (or finite sequences) of length n with respect to the probability distribution P_X and the non-negative number ϵ [21, Ch. 3], [22], i.e., we have

$$\mathcal{T}_\epsilon^n(P_X) = \left\{x^n : \left|\frac{N(a|x^n)}{n} - P_X(a)\right| \le \epsilon P_X(a),\ \forall a \in \mathcal{X}\right\},$$

where $N(a|x^n)$ is the number of occurrences of a in x^n.

1.2.3 Wiretap Channel

Consider the wiretap channel depicted in Fig. 1.1. Alice has a message M that is destined for Bob but should be kept secret from Eve. The message M is uniformly distributed over $\{1,\dots,L\}$, $L = 2^{nR}$, and an encoder $f(\cdot)$ maps M, W to the string

$$X^n = f(M, W) \tag{1.5}$$

with the help of a random variable W that is independent of M and uniformly distributed over $\{1,\dots,L_1\}$, $L_1 = 2^{nR_1}$. The purpose of W is to confuse Eve so that she learns little about M. X^n is transmitted through a memoryless channel $P_{YZ|X}$. Bob observes the channel output Y^n while Eve observes Z^n. The pair MZ^n has the joint distribution P_{MZ^n}. Bob estimates $\hat{M}$ from Y^n and the average error probability is

$$P_{\mathrm{e}}^{(n)} = \Pr\left[\hat{M} \ne M\right]. \tag{1.6}$$

Eve tries to learn M from Z^n and secrecy is measured by

$$\begin{aligned} D(P_{MZ^n}\|P_M Q_{Z^n}) &= \sum_{\substack{(m,z^n)\\ \in \mathrm{supp}(P_{MZ^n})}} P(m,z^n)\log\left(\frac{P(m,z^n)}{P(m)\cdot Q(z^n)}\cdot\frac{P(z^n)}{P(z^n)}\right) \\ &= \sum_{\substack{(m,z^n)\\ \in \mathrm{supp}(P_{MZ^n})}} P(m,z^n)\left(\log\frac{P(z^n|m)}{P(z^n)}+\log\frac{P(z^n)}{Q(z^n)}\right) \\ &= I(M;Z^n)+D(P_{Z^n}\|Q_{Z^n}), \end{aligned} \tag{1.7}$$

where P_{Z^n} is the distribution that Eve observes and Q_{Z^n} is the *default* distribution that Eve expects to observe if Alice is *not* sending useful information.

For example, suppose Alice's default behavior is to transmit x^n with memoryless distribution $Q_X^n(x^n)$. The default output distribution is then

$$Q(z^n) = \sum_{x^n\in\mathrm{supp}(Q_X^n)} Q_X^n(x^n)P_{Z|X}^n(z^n|x^n) = P_Z^n(z^n), \tag{1.8}$$

where $P_Z(z)=\sum_x Q_X(x)P_{Z|X}(z|x)$. When Alice sends useful messages, then P_{Z^n} and Q_{Z^n} are different in general. But if we can make $D(P_{MZ^n}\|P_M Q_{Z^n})$ small then both $I(M;Z^n)$ and $D(P_{Z^n}\|Q_{Z^n})$ are small, which in turn implies (as we shall see) that Eve learns little about M and cannot recognize whether Alice is communicating anything meaningful.

We will consider the case $Q(z^n)=Q_Z^n(z^n)$, i.e., the default behavior has a memoryless distribution. Of course, other distributions may also be interesting. We say that a rate R is *achievable* if for any $\xi_1,\xi_2>0$ there is a sufficiently large n and an encoder and a decoder such that

$$P_{\mathrm{e}}^{(n)} \le \xi_1, \tag{1.9}$$

$$D(P_{MZ^n}\|P_M Q_Z^n) \le \xi_2. \tag{1.10}$$

The *effective secrecy capacity* C_{S} is the supremum of the set of achievable R. We wish to determine C_{S}.

1.3 Effective Secrecy Capacity

We prove the following result:

THEOREM 1.1 *C_{S} is zero if there is no Q_X such that $P_Z=Q_Z$, and otherwise*

$$C_{\mathrm{S}} = \max_{Q_{VX}:\,P_Z=Q_Z} [I(V;Y)-I(V;Z)], \tag{1.11}$$

where the maximization is over all joint distributions Q_{VX} and we have the Markov chain

$$V–X–YZ. \tag{1.12}$$

One may restrict the cardinality of V to $|\mathcal{V}|\le|\mathcal{X}|$.

REMARK 1.1 *Theorem 1.1 applies to random variables with discrete and finite alphabets. However, extensions to real-valued channels such as additive white Gaussian noise (AWGN) channels with power constraints are possible.*

REMARK 1.2 *If one can choose Q_Z to be the P_Z corresponding to the capacity-achieving output distribution of the wiretap channel, then the effective-secrecy capacity is the same as the wiretap channel capacity with (weak or) strong secrecy.*

REMARK 1.3 *Consider a physically degraded channel where X–Y–Z forms a Markov chain. We have (see [2, p. 342])*

$$\begin{aligned} I(V;Y) - I(V;Z) &= I(V;Y|Z) \\ &\le I(X;Y|Z) \\ &= I(X;Y) - I(X;Z), \end{aligned} \tag{1.13}$$

so that choosing $V = X$ achieves capacity.

REMARK 1.4 *The capacity* (1.11) *of a general wiretap channel depends only on the marginals $P(y|x)$ and $P(z|x)$. Hence, the capacity of a (stochastically degraded) channel whose marginals $P(y|x)$ and $P(z|x)$ are the same as those of a physically degraded channel has the same capacity as this physically degraded channel.*

1.3.1 Examples

We consider two examples to show how the stealth requirement impacts C_S. These examples show that fixing $P_Z = Q_Z$ can make calculating C_S rather easy.

EXAMPLE 1.1 *Consider the binary symmetric channels (BSCs)*

$$Y = X \oplus A_1, \quad Z = X \oplus A_2, \tag{1.14}$$

where the alphabet of all random variables is $\{0,1\}$, the operator $\oplus$ is addition modulo 2, A_1 and A_2 are independent of X, $P_{A_1}(1) = p_1$, and $P_{A_2}(1) = p_2$. Suppose that $0 \le p_1 \le p_2 \le 1/2$. If $p_2 = 1/2$ then $I(X;Z) = 0$ and the only interesting case is uniform Q_Z, for which C_S is the same as the BSC capacity. So consider $p_2 < 1/2$ and suppose the default behavior is $Q_Z(1) = q$, where $p_2 \le q \le (1 - p_2)$. We compute

$$P_X(1) = \frac{q - p_2}{1 - 2p_2} \tag{1.15}$$

and, since the channel is stochastically degraded, we have

$$\begin{aligned} C_S = {} & H_b(p_2) - H_b(p_1) - H_b(q) \\ & + H_b\left((q - p_2)\frac{1 - 2p_1}{1 - 2p_2} + p_1\right), \end{aligned} \tag{1.16}$$

where $H_b(\cdot)$ is the binary entropy function. Choosing $q = 1/2$ gives $H(X) = 1$ and we recover the wiretap channel capacity, as expected. But if $q = p_2$ or $q = 1 - p_2$ then $H(X) = 0$ and $C_S = 0$.

EXAMPLE 1.2 *Consider next the AWGN channels*

$$Y = X + A_1, \quad Z = X + A_2 \tag{1.17}$$

with the power constraint $\mathrm{E}[X^2] \le P$. The random variables A_1 and A_2 are independent of X, Gaussian, zero-mean, and have variances N_1 and N_2, respectively. We consider $0 \le N_1 \le N_2$. Suppose the default Z is a Gaussian random variable with zero mean and variance Q, where $N_2 \le Q \le P + N_2$. We thus require that X is zero-mean Gaussian with variance $Q - N_2$. We assume that Theorem 1.1 applies to such channels, and since the channel is stochastically degraded, we compute

$$C_S = \frac{1}{2}\log_2\left(1 + \frac{Q - N_2}{N_1}\right) - \frac{1}{2}\log_2\left(\frac{Q}{N_2}\right), \tag{1.18}$$

where the capacity is measured in bits per channel use. Choosing $Q = P + N_2$ implies $\mathrm{E}[X^2] = P$, and we recover the wiretap channel capacity. But if $Q = N_2$ then $\mathrm{E}[X^2] = 0$ and $C_S = 0$ (this is the regime of covert communication [16–18], see Section 1.2.1). Furthermore, the power required for the default transmissions increases with C_S.

1.3.2 Achievability

We use random coding and the proof technique of [8]. We assume that there is a Q_X for which $P_Z = Q_Z$.

Random code: Fix a distribution Q_X for which $P_Z = Q_Z$ and generate $L \cdot L_1$ codewords $x^n(m,w)$, $m = 1,\dots,L$, $w = 1,\dots,L_1$ using $\prod_{i=1}^n Q_X(x_i(m,w))$. This defines the codebook

$$\mathcal{C} = \{x^n(m,w), m = 1,\dots,L,\ w = 1,\dots,L_1\} \tag{1.19}$$

and we denote the random codebook by

$$\widetilde{\mathcal{C}} = \{X^n(m,w)\}_{(m,w)=(1,1)}^{(L,L_1)}. \tag{1.20}$$

Encoding: To send a message m, Alice chooses w uniformly from $\{1,\dots,L_1\}$ and transmits $x^n(m,w)$. Hence, for any $\mathcal{C}$ we have

$$P_{X^n|\widetilde{\mathcal{C}}}\left(x^n(m,w)\,|\,\mathcal{C}\right) = \frac{1}{L \cdot L_1}. \tag{1.21}$$

Since (1.21) is not $Q_X^n(x^n(m,w))$, the P_{Z^n} is not the desired Q_Z^n in general, see (1.8). Furthermore, we have

$$P_{Z^n|M\widetilde{\mathcal{C}}}(z^n|m,\mathcal{C}) = \sum_{w=1}^{L_1} \frac{1}{L_1} \cdot P_{Z|X}^n(z^n|x^n(m,w)), \tag{1.22}$$

$$P_{Z^n|\widetilde{\mathcal{C}}}(z^n|\mathcal{C}) = \sum_{m=1}^{L}\sum_{w=1}^{L_1} \frac{1}{L \cdot L_1} \cdot P_{Z|X}^n(z^n|x^n(m,w)). \tag{1.23}$$

Bob: Bob knows $\mathcal{C}$ and puts out $(\hat{m},\hat{w})$ if there is a unique pair $(\hat{m},\hat{w})$ satisfying the typicality check

$$(x^n(\hat{m},\hat{w}),y^n) \in \mathcal{T}_\epsilon^n(Q_X P_{Y|X}). \tag{1.24}$$

Otherwise he puts out $(\hat{m},\hat{w}) = (1,1)$.

Analysis: Define the real-valued random variables

$$E_1 = \Pr\left[(\hat{M},\hat{W}) \neq (M,W)|\widetilde{\mathcal{C}}\right], \tag{1.25}$$

$$E_2 = D\left(P_{MZ^n|\widetilde{\mathcal{C}}} \,\middle\|\, P_M Q_Z^n\right). \tag{1.26}$$

E_1 is Bob's block decoding error probability, and E_2 represents the security (secrecy and stealth) with respect to Eve's receiver. Using standard arguments (see [22]), $\mathrm{E}[E_1]$ can be made small with large n as long as

$$R + R_1 < I(X;Y). \tag{1.27}$$

To bound $\mathrm{E}[E_2]$, we use the steps in [8, Eq. (9)] to write

$$\begin{aligned}
&\mathrm{E}\left[D\left(P_{MZ^n|\widetilde{\mathcal{C}}} \,\middle\|\, P_M Q_Z^n\right)\right] \\
&\overset{(a)}{=} \mathrm{E}\left[\log \frac{\sum_{j=1}^{L_1} P_{Z|X}^n(Z^n|X^n(M,j))}{L_1 \cdot Q_Z^n(Z^n)}\right] \\
&= \sum_{m=1}^{L}\sum_{w=1}^{L_1} \frac{1}{L\cdot L_1} \mathrm{E}\left[\log \frac{\sum_{j=1}^{L_1} P_{Z|X}^n(Z^n|X^n(m,j))}{L_1 \cdot Q_Z^n(Z^n)} \middle| M=m, W=w\right] \\
&\overset{(b)}{\leq} \sum_{m=1}^{L}\sum_{w=1}^{L_1} \frac{1}{L\cdot L_1} \mathrm{E}\left[\log\left(\frac{P_{Z|X}^n(Z^n|X^n(m,w))}{L_1 \cdot Q_Z^n(Z^n)} + 1\right) \middle| M=m, W=w\right] \\
&\overset{(c)}{=} \mathrm{E}\left[\log\left(\frac{P_{Z|X}^n(Z^n|X^n)}{L_1 \cdot Q_Z^n(Z^n)} + 1\right)\right],
\end{aligned} \tag{1.28}$$

where

(a) follows from (1.22) and by taking the expectation over M, W, $X^n(1,1),\ldots,X^n(L,L_1)$, Z^n;
(b) follows by applying Jensen's inequality to the expectation over the $X^n(m,j), j \neq w$ for a fixed m, using (1.8), and using $P_Z = Q_Z$;
(c) follows by choosing $X^n Z^n \sim Q_X^n P_{Z|X}^n$.

We may write (1.28) as

$$\mathrm{E}\left[\log\left(\frac{P_{Z|X}^n(Z^n|X^n)}{L_1 \cdot Q_Z^n(Z^n)} + 1\right)\right] = d_1 + d_2, \tag{1.29}$$

where the expectation is split based on typical pairs via

$$d_1 = \sum_{(x^n,z^n)\in\mathcal{T}_\epsilon^n(Q_XP_{Z|X})} Q_X^n(x^n)P_{Z|X}^n(z^n|x^n)\log\left(\frac{P_{Z|X}^n(z^n|x^n)}{L_1\cdot Q_Z^n(z^n)}+1\right),$$

$$d_2 = \sum_{\substack{(x^n,z^n)\notin\mathcal{T}_\epsilon^n(Q_XP_{Z|X})\\(x^n,z^n)\in\mathrm{supp}(Q_X^nP_{Z|X}^n)}} Q_X^n(x^n)P_{Z|X}^n(z^n|x^n)\log\left(\frac{P_{Z|X}^n(z^n|x^n)}{L_1\cdot Q_Z^n(z^n)}+1\right).$$

Using standard inequalities (see [22, Lemmas 18 and 20]) we have

$$\begin{aligned} d_1 &\le \log\left[\frac{2^{-n(1-\epsilon)H(Z|X)}}{L_1\cdot 2^{-n(1+\epsilon)[H(Z)+D(P_Z\|Q_Z)]}}+1\right]\\ &\overset{(a)}{=} \log\left[2^{-n(R_1-I(X;Z)-\kappa\epsilon)}+1\right]\\ &\le \log(e)\cdot 2^{-n(R_1-I(X;Z)-\kappa\epsilon)}, \end{aligned} \tag{1.30}$$

where (a) follows because we chose Q_X so that $P_Z = Q_Z$, and κ is a constant independent of n. We find that $d_1 \to 0$ if $n\to\infty$ and

$$R_1 > I(X;Z)+\kappa\epsilon. \tag{1.31}$$

Next, consider d_2 and a pair (x^n,z^n) in the support of $Q_X^nP_{Z|X}^n$. We have

$$Q_Z^n(z^n) = P_Z^n(z^n) = \sum_{\tilde{x}^n\in\mathrm{supp}(Q_X^n)} Q_X^n(\tilde{x}^n)P_{Z|X}^n(z^n|\tilde{x}^n), \tag{1.32}$$

so that $Q_Z^n(z^n)$ is positive. We bound (see [22, Lemma 17])

$$\begin{aligned} d_2 &\le \sum_{\substack{(x^n,z^n)\notin\mathcal{T}_\epsilon^n(Q_XP_{Z|X})\\(x^n,z^n)\in\mathrm{supp}(Q_X^nP_{Z|X}^n)}} Q_X^n(x^n)P_{Z|X}^n(z^n|x^n)\log\left[\left(\frac{1}{\mu_Z}\right)^n+1\right]\\ &\le 2|\mathcal{X}|\cdot|\mathcal{Z}|\cdot e^{-n\epsilon^2\mu_{XZ}/3}\log\left[\left(\frac{1}{\mu_Z}\right)^n+1\right], \end{aligned} \tag{1.33}$$

where

$$\mu_Z = \min_{z\in\mathrm{supp}(Q_Z)}Q(z), \tag{1.34}$$

$$\mu_{XZ} = \min_{(x,z)\in\mathrm{supp}(Q_XP_{Z|X})}Q(x)P(z|x). \tag{1.35}$$

If $\frac{1}{\mu_Z}<1$, we have

$$d_2 \le 2|\mathcal{X}|\cdot|\mathcal{Z}|\cdot e^{-n\epsilon^2\mu_{XZ}/3}\cdot\log 2 \tag{1.36}$$

and $d_2\to 0$ as $n\to\infty$. If $\frac{1}{\mu_Z}\ge 1$, we have

$$d_2 \le 2|\mathcal{X}|\cdot|\mathcal{Z}|\cdot e^{-n\epsilon^2\mu_{XZ}/3}\cdot n\cdot\log\left(\frac{1}{\mu_Z}+1\right) \tag{1.37}$$

and $d_2\to 0$ as $n\to\infty$.

For any $\xi_1, \xi_2 > 0$, define the event

$$\mathcal{E} = \{E_1 > \xi_1 \text{ or } E_2 > \xi_2\}. \tag{1.38}$$

Using the union bound and Markov's inequality, we obtain

$$\begin{aligned} \Pr\left[\mathcal{E}\right] &\leq \Pr[E_1 > \xi_1] + \Pr[E_2 > \xi_2] \\ &\leq \frac{1}{\xi_1}\mathrm{E}[E_1] + \frac{1}{\xi_2}\mathrm{E}[E_2]. \end{aligned} \tag{1.39}$$

Combining the above, we can make $\Pr[\mathcal{E}] \to 0$ as $n \to \infty$ as long as

$$R + R_1 < I(X;Y), \tag{1.40}$$

$$R_1 > I(X;Z). \tag{1.41}$$

We hence have the achievability of any R satisfying

$$0 \leq R < \max_{Q_X : P_Z = Q_Z} [I(X;Y) - I(X;Z)]. \tag{1.42}$$

Of course, if the right-hand side of (1.42) is non-positive, then we require $R = 0$.

Finally, following [2] we prefix a channel $Q_{X|V}$ to the channel $P_{YZ|X}$ and obtain a new channel $Q_{YZ|V}$ where

$$Q(y,z|v) = \sum_x Q(x|v) P(y,z|x). \tag{1.43}$$

Using a similar analysis as above, we have the achievability of any R satisfying

$$0 \leq R < \max_{Q_{VX} : P_Z = Q_Z} [I(V;Y) - I(V;Z)], \tag{1.44}$$

where the maximization is over all Q_{VX} satisfying (1.12). Again, if the right-hand side of (1.44) is non-positive, then we require $R = 0$. As usual, the purpose of adding the auxiliary variable V is to potentially increase R. Note that $V = X$ recovers (1.42). Hence, the right-hand side of (1.42) is at most the right-hand side of (1.44).

REMARK 1.5 *The average divergence* $\mathrm{E}[D(P_{MZ^n|\widetilde{\mathcal{C}}} || P_M Q_Z^n)]$ *is the sum of* $I(M\widetilde{\mathcal{C}}; Z^n)$ *and* $D(P_{Z^n} || Q_Z^n)$ *[6, Section III] (see also [8, Section III-B]). To see this, consider*

$$\begin{aligned} &\mathrm{E}[D(P_{MZ^n|\widetilde{\mathcal{C}}} || P_M Q_Z^n)] \\ &= D(P_{MZ^n|\widetilde{\mathcal{C}}} || P_M Q_Z^n | P_{\widetilde{\mathcal{C}}}) \\ &\overset{(a)}{=} D(P_{Z^n|M\widetilde{\mathcal{C}}} || Q_Z^n | P_M P_{\widetilde{\mathcal{C}}}) \\ &= D(P_{Z^n|M\widetilde{\mathcal{C}}} || P_{Z^n|M} P_{\widetilde{\mathcal{C}}}) + D(P_{Z^n} || Q_Z^n) \\ &= I(M\widetilde{\mathcal{C}}; Z^n) + D(P_{Z^n} || Q_Z^n), \end{aligned} \tag{1.45}$$

where (a) follows by the independence of M *and the codewords. Therefore, as* $\mathrm{E}[D(P_{MZ^n|\widetilde{\mathcal{C}}} || P_M Q_Z^n)] \to 0$ *we have* $I(M\widetilde{\mathcal{C}}; Z^n) \to 0$*, which means that* $M\widetilde{\mathcal{C}}$ *and* Z^n *are (almost) independent. This makes sense, since for effective secrecy the adversary learns little about* M *and the presence of meaningful transmission.*