Channel Upgrading for Semantically-Secure Encryption on Wiretap Channels

Ido Tal
Department of Electrical Engineering
Techion, Haifa 32000, Israel
idotal@ieee.org

Alexander Vardy
University of California San Diego
La Jolla, CA 92093, USA
avardy@ucsd.edu

***Abstract*—Bellare and Tessaro recently introduced a new coding scheme, based on cryptographic principles, that guarantees strong security for a wide range of symmetric wiretap channels. This scheme has numerous advantages over alternative constructions, including constructions based on polar codes. However, it achieves secrecy capacity only under a certain restrictive condition. Specifically, let V be the main channel (from Alice to Bob) and let W be wiretap channel (from Alice to Eve). Suppose that W has a finite output alphabet $\mathcal{Y}$, and let X and Y denote the input and output of W, respectively. Then the rate of the Bellare-Tessaro coding scheme is at most $I(V) - \Psi(W)$, where $I(V)$ is the capacity of V and $\Psi(W)$ is given by**

$$\Psi(W) \overset{\text{def}}{=} \log_2 |\mathcal{Y}| - H(\boldsymbol{Y}|\boldsymbol{X})$$

For symmetric channels, it is clear that $\Psi(W) \geqslant I(W)$ with equality if and only if uniform input to W produces uniform output. Unfortunately, few symmetric DMCs satisfy this condition.

In this paper, we show how the Bellare-Tessaro coding scheme can be extended to achieve secrecy capacity in the case where W is an *arbitrary* symmetric DMC. To this end, we solve the following problem. Given W and $\varepsilon > 0$, we construct another channel Q such that W is degraded with respect to Q while the difference between $\Psi(Q)$ and $I(W)$ is at most ε. We also solve a closely related problem, where the output alphabet of Q is required to be of a given size M. In this case, we construct a channel Q that is *equivalent* to W, such that $\Psi(Q)$ is a small as possible. We furthermore extend these results, and thereby the applicability of the Bellare-Tessaro coding scheme, to channels with binary input and *continuous output*.

I. Introduction

We consider the classical wiretap channel model, introduced by Wyner [12] in 1975. In this setting, Alice wishes to send messages to Bob through a communication channel V, called the ***main channel***, but her transmissions also reach an adversary Eve through another channel W, called the ***wiretap channel***. The goal is to design a coding scheme — namely, an encoding algorithm and a decoding algorithm — that makes it possible to communicate both *reliably* and *securely*. The highest rate at which such communication is possible is known as the ***secrecy capacity*** and denoted $\mathcal{C}_s$. In the case where the channels V and W are both symmetric (as we henceforth assume) and W is degraded with respect to V, the secrecy capacity is given by

$$\mathcal{C}_s = I(V) - I(W) \tag{1}$$

where $I(\cdot)$ denotes channel capacity. In his seminal paper [12], Wyner showed that both reliability and (weak) security can be achieved using random coding at rates approaching the secrecy capacity. However, the problem of achieving the secrecy capacity *explicitly and with low complexity* remained open. In the past three years, a solution to this problem based on polar coding [1], for the general class of symmetric and degraded wiretap channels, was presented in [9,10] and other papers.

More recently, a very different approach to this problem was proposed in the papers of Bellare and Tessaro [2] and Bellare, Tessaro, and Vardy [4]. The Bellare-Tessaro coding scheme of [2,4] is based upon cryptographic principles, in particular the notion of *invertible extractors*. It is shown in [2,4] that multiplication over a finite field provides an efficient instantiation of an invertible extractor. Thus the encoding algorithm of [2, 4] proceeds in three simple steps as follows. The first step consists of taking an m-bit message $\boldsymbol{M} \in \{0,1\}^m$ and appending to it r bits chosen uniformly at random from $\{0,1\}^r$ to generate a vector $\boldsymbol{V}$ with $k = m + r$ bits. In the second step, the vector $\boldsymbol{V}$ is regarded as an element of the finite field $\mathrm{GF}(2^k)$. This step employs a k-bit uniformly random seed $\boldsymbol{S}$, which is assumed to be known *a priori* to all parties. The second step consists of multiplying $\boldsymbol{V}$ and $\boldsymbol{S}$ in $\mathrm{GF}(2^k)$ to generate a k-bit vector $\boldsymbol{U}$. The third and final step consists of encoding $\boldsymbol{U}$ with a binary linear (n,k) code $\mathbb{C}$ to generate an n-bit vector $\boldsymbol{X}$. The vector $\boldsymbol{X}$ serves as the input to both V and W (if the input alphabet of V and W is non-binary, any bijective mapping from bits to symbols can be used).

We shall refer to the encoding algorithm of the foregoing paragraph as the ***BT coding scheme***. It is shown in [2,4] that the security of this coding scheme *does not depend* on the choice of the error-correcting code $\mathbb{C}$, as long as its rate k/n exceeds the capacity of W. This modularity is one of the key advantages of the BT coding scheme: system designers can use arbitrary error-correction and modulation methods at the output of $\mathrm{GF}(2^k)$-multiplication. Observe that the overall rate of the BT coding scheme is given by $m/n = k/n - r/n$. One can view the terms $1-k/n$ and r/n as "penalties in rate" due to reliability and security requirements, respectively. In order to achieve the secrecy capacity in (1), it is necessary and sufficient to have

$$\lim_{n\to\infty} \frac{k}{n} = I(V) \qquad \text{and} \qquad \lim_{n\to\infty} \frac{r}{n} = I(W) \tag{2}$$

Another key advantage of the BT coding scheme is that it directly provides *semantic security*. Semantic security, first introduced in [8], is now widely regarded as the gold standard in cryptography. As shown by Bellare, Tessaro, and Vardy [3, 4], it is a stronger notion of security than conventional metrics based on equivocation rate or mutual information. Briefly, given a coding scheme $\mathcal{E}$ and a wiretap channel W, the ***semantic-security advantage*** $\mathbf{Adv}_{\mathrm{S}}(\mathcal{E}, W)$ is defined [3,4] as follows:

$$\mathbf{Adv}_{\mathrm{S}}(\mathcal{E},W) \overset{\text{def}}{=} \sup_{P_M,f}\left(\sup_{\mathcal{A}}\Pr\left\{\mathcal{A}(\boldsymbol{Y}) = f(\boldsymbol{M})\right\} - \sup_{\mathcal{S}}\Pr\left\{\mathcal{S} = f(\boldsymbol{M})\right\}\right)$$

Here, f is an arbitrary function of the message $\boldsymbol{M}$, and $\sup_{P_{\boldsymbol{M}},f}$ maximizes over *all possible* functions f and message distributions $P_{\boldsymbol{M}}$. Further, $\boldsymbol{Y}$ is the vector at the output of the wiretap channel W, and $\sup_{\mathcal{A}}$ maximizes over all possible adversary strategies $\mathcal{A}$. Thus the term within the parentheses is the maximum probability that an adversary, given its observations $\boldsymbol{Y}$, can compute the value of the function f on the message, minus the maximum probability that a simulator $\mathcal{S}$ can do the same given no observations at all. In cryptography, it is customary to measure security in bits: we say that a coding scheme $\mathcal{E}$ provides σ ***bits of semantic security*** on the wiretap channel W if $\mathbf{Adv}_{\mathrm{S}}(\mathcal{E}, W) \leqslant 2^{-\sigma}$. This definition gives a quantitative, rather than qualitative, measure of security. For more details on the definitions in this paragraph, see [3,4,8].

In the next paragraph, we establish the relationship between the number of random bits r required by the BT coding scheme and the number of bits of semantic security it provides. But first, we need some notation. We henceforth assume that the wiretap channel $W\colon \mathcal{X} \to \mathcal{Y}$ is a symmetric, discrete, memoryless channel. For such channels, we define:

$$W(y) \overset{\text{def}}{=} \frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}} W(y|x) \tag{3}$$

$$H(W) \overset{\text{def}}{=} -\sum_{y\in\mathcal{Y}} W(y|x)\log_2 W(y|x)\,, \quad \text{for } x\in\mathcal{X} \tag{4}$$

$$\Psi(W) \overset{\text{def}}{=} \log_2|\mathcal{Y}| - H(W) \tag{5}$$

Note that the definition of $H(W)$ in (4) does not depend on the choice of $x\in\mathcal{X}$, since the rows of the transition probability matrix of W are permutations of each other. Also note that the capacity $I(W)$ can be expressed as the difference between the entropy of $W(\cdot)$ in (3) and $H(W)$. Therefore $\Psi(W) \geqslant I(W)$, with equality iff $W(\cdot)$ is the uniform distribution on $\mathcal{Y}$.

The following proposition follows from the results of [2,4]; we omit the proof.

Proposition 1. *Let $W\colon \mathcal{X} \to \mathcal{Y}$ be a symmetric, discrete, memoryless channel. Then the BT coding scheme $\mathcal{E}_{n,k,r}$ with parameters $n, k,$ and r guarantees σ bits of semantic security on the wiretap channel W, provided that*

$$r = \left\lceil 2(\sigma+1) + \sqrt{n}\log_2(|\mathcal{Y}|+3)\sqrt{2(\sigma+3)} + n\Psi(W)\right\rceil$$

It follows from Proposition 1 that $\lim_{n\to\infty} r/n = \Psi(W)$. In light of (2) and the observations in the foregoing paragraph, we see that the BT coding scheme achieves secrecy capacity if and only if uniform input to W induces uniform output. This was, indeed, also observed in [2,4]. While this property holds for binary symmetric channels (BSC), symmetric channels with a large output alphabet rarely satisfy this condition.

In this paper, we show how the BT coding scheme can be extended to achieve secrecy capacity in the case where the wiretap channel W is an *arbitrary* symmetric DMC. To this end, we first solve the following problem. Given W and $\varepsilon > 0$, we show how to construct another channel Q such that W *is degraded with respect to Q and $\Psi(Q) - I(W) \leqslant \varepsilon$*. The key point is this: any coding scheme that is secure for Q is necessarily secure for W. To see this, note that one possible adversary strategy for Q is to start by degrading Q into W. Thus, given the wiretap channel W, we first construct Q and then design the BT coding scheme for Q rather than W, so that

$$\lim_{n\to\infty}(r/n) = \Psi(Q) \leqslant I(W) + \varepsilon$$

The rest of this paper is organized as follows. In Section II, we define the concept of an *equivalent channel*, and then show how any symmetric, discrete, memoryless channel W can be replaced by an equivalent channel Q for which $\Psi(Q)$ is arbitrarily close to $I(W)$. This suffices to establish our asymptotic results (cf. Theorem 4). However, for finite n, we also care about the coefficient of $\sqrt{n}$ in Proposition 1. This coefficient is roughly proportional to $\log_2 M$, where M is the size of the output alphabet of Q. Thus we consider the situation where the output alphabet of Q is constrained to be of a given size. In Section III, we present an algorithm that, given W and M, produces an equivalent channel $Q : \mathcal{X} \to \mathcal{Z}$ such that $|\mathcal{Z}| = M$ and $\Psi(Q)$ is as small as possible. Finally, in Section IV, we extend our results to the case where W has binary input but continuous output (e.g. the binary-input AWGN channel).

II. Symmetric Channels and Letter Splitting

In this section, we first set up the basic notation and concepts. We then prove the claim that a fraction of only $I(W)$ random bits asymptotically suffices for security.

A. Channels

A discrete memoryless channel (DMC) W with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ will be denoted as $W : \mathcal{X} \to \mathcal{Y}$. We say that the DMC W is ***symmetric (SDMC)*** if [7, Page 94] the set of output symbols $\mathcal{Y}$ can be partitioned into subsets $\mathcal{Y}_1, \mathcal{Y}_2, \ldots, \mathcal{Y}_T$ such that the following holds. For a each such subset $\mathcal{Y}_t$, let A_t be the matrix for which the rows are indexed by $\mathcal{X}$, the columns by $\mathcal{Y}_t$, and for which entry $(x, y) \in \mathcal{X}\times\mathcal{Y}_t$ is equal to $W(y|x)$. Then rows of A_t are permutation of each other and column of A_t are permutations of each other.

B. Degraded, Upgraded, and Equivalent Channels

A DMC $W : \mathcal{X} \to \mathcal{Y}$ is (stochastically) ***degraded*** with respect to a DMC $Q : \mathcal{X} \to \mathcal{Z}$, denoted $W \preceq Q$, if there exists an intermediate channel $P : \mathcal{Z} \to \mathcal{Y}$ such that

$$W(y|x) = \sum_{z\in\mathcal{Z}} Q(z|x)\cdot P(y|z)\ .$$

Namely, W is the result of concatenating the channels Q and P. Alternatively, we say that Q is ***upgraded*** with respect to W, and denote this as $Q \succeq W$. If W is both upgraded and degraded with respect to Q, then we say that W and Q are ***equivalent***, and denote this as $Q \equiv W$. It follows from the data-processing inequality [6, Theorem 2.8.1] that $Q \succeq W$ implies $I(Q) \geqslant I(W)$. Thus, $Q \equiv W$ implies $I(Q) = I(W)$.

Corollary 2. *Let $W : \mathcal{X} \to \mathcal{Y}$, σ, r, and n be as in Proposition 1. Suppose the SDMC $Q : \mathcal{X} \to \mathcal{Z}$ is upgraded with respect to W. Then, we can substitute W by Q in Proposition 1. That is, we can take*

$$r = 2(\sigma+1) + \sqrt{n}\log_2(|\mathcal{Z}|+3)\sqrt{2(\sigma+3)} + n \cdot \Psi(Q) \ . \quad (6)$$

C. Letter Splitting

Let $W : \mathcal{X} \to \mathcal{Y}$ be a given SDMC with a corresponding partition $\mathcal{Y}_1, \mathcal{Y}_2, \ldots, \mathcal{Y}_T$. We say that a function $s : \mathcal{Y} \to \mathbb{N}$ from $\mathcal{Y}$ to the positive integers is an *output letter split* of W if for all $1 \leqslant t \leqslant T$ and all $y, y' \in \mathcal{Y}_t$ we have that $s(y) = s(y')$. Thus, the split function s assigns the same value to every letter in a subset $\mathcal{Y}_t$, and we can thus abuse notation and define $s(\mathcal{Y}_t)$ as that value. The SDMC corresponding to this split, $Q : \mathcal{X} \to \mathcal{Z}$ is defined as follows. The output alphabet of Q is gotten by duplicating each letter $y \in \mathcal{Y}$ and making $s(y)$ distinct copies:

$$\mathcal{Z} = \bigcup_{y \in \mathcal{Y}} \{y_1, y_2, \ldots, y_s \mid s = s(y)\} \ .$$

The transition probabilities of Q are given by

$$Q(y_i|x) = W(y|x)/s(y) \ , \quad (x, y_i) \in \mathcal{X} \times \mathcal{Z} \ .$$

Namely, each letter y is duplicated $s(y)$ times, and the conditional probability of receiving each copy is simply $1/s(y)$ times the corresponding probability in the original channel W.

Note that since W is a SDMC, then so is Q. Also, note that W and Q are equivalent channels, $W \equiv Q$.

Let an SDMC $W : \mathcal{X} \to \mathcal{Y}$ be given. As previously discussed, a discrepancy between $I(W)$ and $\Psi(W)$ arises when a uniform distribution on the input to W does not result in a uniform distribution on the output. We will now use the above mentioned splitting operation to define a channel $Q : \mathcal{X} \to \mathcal{Z}$. The merit of Q will be that a uniform input distribution results in an output distribution that is close to uniform. The price we will pay for this quasi-uniformity is a larger output alphabet, compared to that of W (recall that the coefficient of $\sqrt{n}$ in (6) is an increasing function of the output alphabet size).

Lemma 3. *Let $W : \mathcal{X} \to \mathcal{Y}$ be an SDMC, and let $M \geqslant 1$ be a given positive integer. For each $y \in \mathcal{Y}$, define*

$$s(y) = \lceil M \cdot W(y) \rceil \ ,$$

and let $Q : \mathcal{X} \to \mathcal{Z}$ be the channel resulting from the applying the letter splitting function s to W. Then,

$$\Psi(Q) - I(W) = \Psi(Q) - I(Q) \leqslant \log_2\left(1 + \frac{|\mathcal{Y}|}{M}\right) \ , \quad (7)$$

and

$$|\mathcal{Z}| \leqslant M + |\mathcal{Y}| \ . \quad (8)$$

Proof: The bound (8) follows by

$$|\mathcal{Z}| = \sum_{y \in \mathcal{Y}} s(y) = \sum_{y \in \mathcal{Y}} \lceil M \cdot W(y) \rceil \leqslant \sum_{y \in \mathcal{Y}} 1 + M \cdot W(y) = |\mathcal{Y}| + M \ .$$

We now prove (7). We start by simplifying $\Psi(Q) - I(Q)$ to

$$\Psi(Q) - I(Q) = \log_2 |\mathcal{Z}| + \sum_{z \in \mathcal{Z}} Q(z) \log_2 Q(z) = \log_2 |\mathcal{Z}| + \sum_{y \in \mathcal{Y}} W(y) \log_2 \left(\frac{W(y)}{s(y)}\right) \ .$$

We have already proved (8), and thus have an upper bound on the first term. Thus, we concentrate now on the second term.

$$\sum_{y \in \mathcal{Y}} W(y) \log_2 \left(\frac{W(y)}{s(y)}\right) = \sum_{y \in \mathcal{Y}} W(y) \log_2 \left(\frac{W(y)}{\lceil M \cdot W(y) \rceil}\right) \leqslant \sum_{y \in \mathcal{Y}} W(y) \log_2 \left(\frac{W(y)}{M \cdot W(y)}\right) = -\log_2 M \ .$$

Combining the above two bounds, we get

$$\Psi(Q) - I(Q) \leqslant \log_2(|\mathcal{Y}| + M) - \log_2 M = \log_2\left(1 + \frac{|\mathcal{Y}|}{M}\right) \ .$$

■

Combining Corollary 2 with Lemma 3 gives us the following theorem.

Theorem 4. *Let $W : \mathcal{X} \to \mathcal{Y}$, σ, r, and n be as in Proposition 1. Let $M \geqslant 1$ be a parameter we are allowed to choose. Then, the number of random bits needed to achieve semantic security is at most*

$$r = 2(\sigma + 1) + \sqrt{n}\log_2(M + |\mathcal{Y}| + 3)\sqrt{2(\sigma+3)} + n \cdot \left(I(W) + \log_2\left(1 + \frac{|\mathcal{Y}|}{M}\right)\right) \ . \quad (9)$$

Setting, say, $M = n$ gives $\lim_{n\to\infty}(r/n) = I(W)$.

III. Optimal Letter Splitting

Recall that Theorem 4 arises from choosing a specific letter-splitting function, $s(y) = \lceil n \cdot W(y) \rceil$. Although the theorem states that this choice is a good choice asymptotically, a natural direction to pursue now is the finite n case. That is, given W, σ, and n, we may ask what is the best letter-splitting function one can choose so that (6) is minimized. We do not know how to answer this question. However, let us pose a related one. Namely, suppose $W : \mathcal{X} \to \mathcal{Y}$ is such that the subsets $\mathcal{Y}_1, \mathcal{Y}_2, \ldots, \mathcal{Y}_T$ are all of the same size, μ. For example, if $|\mathcal{X}| = 2$, then our assumption holds when $\mathcal{Y}$ does not contain an erasure symbol, in which case we can always find a partition for which $\mu = 2$. We now ask, suppose we are given a parameter M which is a multiple of μ, and wish to find a letter-splitting function s for which $\sum_{y \in \mathcal{Y}} s(y) = M$ and for which the resulting channel Q has a minimum $\Psi(Q)$ value.

Algorithm 1: Construction of optimal splitting function

```
Input: Channel $W\colon \mathcal{X} \to \mathcal{Y}$, a partition $\mathcal{Y}_1, \mathcal{Y}_2, \ldots, \mathcal{Y}_T$
       where each subset is of size $\mu$, a positive
       integer $M$ which is a multiple of $\mu$
Output: A letter-splitting function $s$ such that
        $\sum_{y \in \mathcal{Y}} s(y) = M$ and $\Psi(Q)$ is minimal
// Initialization
$s(\mathcal{Y}_1) = s(\mathcal{Y}_2) = \cdots = s(\mathcal{Y}_T) = 1$ ;
// Main loop
for $i = 1, 2, \ldots, \frac{M}{\mu} - T$ do
    $t = \arg\max_{1 \leqslant t \leqslant T} \sum_{y \in \mathcal{Y}_t} W(y) \log_2 \left( \frac{s(\mathcal{Y}_t)+1}{s(\mathcal{Y}_t)} \right)$ ;
    $s(\mathcal{Y}_t) = s(\mathcal{Y}_t) + 1$;
return $s$;
```

We now show that a greedy algorithm can find such a letter-splitting function efficiently.

Theorem 5. *Given a valid input to Algorithm* 1*, the output is a valid letter-splitting function* s*, such that* $\sum_{y \in \mathcal{Y}} s(y) = M$ *and the resulting channel* Q *is such that* $\Psi(Q)$ *is minimized.*

Proof: First, note that after the initialization step, we have $\sum_{y \in \mathcal{Y}} s(y) = \mu \cdot T$. Each iteration obviously increments the sum by μ, so at the end we indeed have a letter-splitting function s such that $\sum_{y \in \mathcal{Y}} s(y) = M$.

With respect to optimality, first note that a channel Q which results from a splitting function has $I(Q) = I(W)$, since $Q \equiv W$. Thus, minimizing $\Psi(Q)$ is equivalent to maximizing

$$I(Q) - \Psi(Q) = \sum_{y \in Y} -W(y) \log_2 \left(\frac{W(y)}{s(y)} \right) - \log_2 M \ .$$

Clearing away constant terms, our target function becomes $\sum_{y \in \mathcal{Y}} W(y) \log_2 s(y)$. Recall that we must have $s(y) \geqslant 1$ for all $y \in \mathcal{Y}$. We now rephrase our optimization problem in an equivalent manner. Define the set

$$A = \bigcup_{y \in \mathcal{Y}} \bigcup_{i=1}^{M/\mu - T} \left\{ \delta(y,i) = W(y) \log_2 \left(\frac{i+1}{i} \right) \right\} \ .$$

Then, finding the optimal $s(y)$ is equivalent to choosing $M/\mu - T$ numbers from the set A such that their sum is maximal, and they satisfy the following constraint: If $\delta(y,i)$ was picked and $i > 1$, then $\delta(y, i-1)$ must be picked as well. To see the equivalence, define $s(y)$ as the largest i such that $\delta(y,i)$ was picked, or as 1 if no such δ was picked. The important point to note now is that the last constraint is redundant. That is, note that since $\log_2$ is a concave function, we have a "diminishing returns" effect $\log_2(i+2) - \log_2(i+1) < \log_2(i+1) - \log_2(i)$. Thus, the optimal solution will satisfy the constraint. Therefore, we can forget about the constraint and simply pick the M largest elements of A. It should now be easy to see that that is exactly what Algorithm 1 does. ■

We note that the complexity of Algorithm 1 is $O\big(\log(T) \cdot (M/\mu - T)\big)$, provided that one uses a heap [5, Chapter 6].

IV. Optimization for Continuous Alphabets

Recall that Theorem 4 has given us a method by which, asymptotically, only a fraction $r/n = I(W)$ random bits need to be utilized in order to achieve semantic security. However, the underlying assumption was that that the channel W had a finite output alphabet size. Specifically, Theorem 4 does not apply if $W : \mathcal{X} \to \mathcal{Y}$ is the binary-input Gaussian channel (BAWGN). More so, $\Psi(W)$ is undefined in this case, so we cannot even fall back to relying on Proposition 1. However, one need not go to such extremes. Even if W does have a finite output alphabet, but that alphabet size is rather large, we stand to lose much as implied by the coefficient of $\sqrt{n}$ in (9). Luckily, if W is a channel with binary input, we can derive a bound on r which is not a function of the output alphabet size of W. Thus, in this section, we will assume that $|\mathcal{X}| = 2$.

Apart from our assumption on a binary input, symmetry (yet to be defined in a non-DMC setting), and memorylessness, we make the following assumption for simplicity of exposition. Let the input alphabet of W be $\mathcal{X} = \{-1, 1\}$ and let the output alphabet be $\mathcal{Y} = \mathbb{R}$ the real numbers. Let the p.d.f. of W be f, and let it be symmetric:

$$f(y|1) = f(-y|-1) \ , \quad y \in \mathbb{R} \ . \tag{10}$$

Next, we assume that

$$f(y|1) \geqslant f(y|-1) \ , \quad y \geqslant 0 \ , \tag{11}$$

and also that the likelihood ratios increase with y. That is, for $y_1 < y_2$ we have

$$\frac{f(y_1|1)}{f(y_1|-1)} \leqslant \frac{f(y_2|1)}{f(y_2|-1)} \ , \quad -\infty < y_1 < y_2 < \infty \ . \tag{12}$$

The above implicitly assumes that $f(y|x) > 0$ for every (x, y). Note that the above conditions hold for the BAWGN channel.

We start by recursively defining the following equi-probable regions. Denote $y_0 = 0$. Next, for $1 \leqslant i < M$ and an already calculated y_{i-1}, let $y_i > y_{i-1}$ be such that

$$\int_{-y_i}^{-y_{i-1}} f(y|1)\, dy + \int_{y_{i-1}}^{y_i} f(y|1)\, dy = \frac{1}{M} \ .$$

Lastly, by abuse of notation, we "define" $y_M = \infty$. Thus, for $1 \leqslant i \leqslant M$, the regions

$$A_i = \{y \ : \ -y_i < y \leqslant -y_{i-1}\} \cup \{y \ : \ y_{i-1} \leqslant y < y_i\}$$

form a partition of $\mathcal{Y} = \mathbb{R}$ and are equi-probable with respect to $f(y|1)$ as well as $f(y|-1)$, by the symmetry condition (10).

For future reference, let us define for $0 \leqslant i \leqslant M$ the sets

$$B_i = \{y \ : \ y_{i-1} \leqslant y < y_i\}$$

Next, for $1 \leqslant i < M$, define

$$\lambda_i = \frac{f(y_i|1)}{f(y_1|-1)} \ ,$$

and let $\lambda_M = \infty$. Then, by (12) and (11), we have for $1 \leqslant i \leqslant M$ that

$$1 \leqslant \lambda_{i-1} = \inf_{y \in B_i} \frac{f(y|1)}{f(y|-1)} \leqslant \sup_{y \in B_i} \frac{f(y|1)}{f(y|-1)} \leqslant \lambda_i \; . \qquad (13)$$

We now define the upgraded channel $Q : \mathcal{X} \to \mathcal{Z}$ and $Q' : \mathcal{X} \to \mathcal{Z}$. The output alphabet of Q is

$$\mathcal{Z} = \{z_1, \bar{z}_1, z_2, \bar{z}_2, \ldots, z_M, \bar{z}_M\} \; . \qquad (14)$$

The channel Q is defined as

$$Q(z|1) = \begin{cases} \frac{\lambda_i}{M(\lambda_i+1)} & \text{if } z = z_i \text{ and } \lambda_i \neq \infty \; , \\ \frac{1}{M(\lambda_i+1)} & \text{if } z = \bar{z}_i \text{ and } \lambda_i \neq \infty \; , \\ \frac{1}{M} & \text{if } z = z_i \text{ and } \lambda_i = \infty \; , \\ 0 & \text{if } z = \bar{z}_i \text{ and } \lambda_i = \infty \; , \end{cases} \qquad (15)$$

and

$$Q(z_i|-1) = Q(\bar{z}_i|1) \; , \quad Q(\bar{z}_i|-1) = Q(z_i|1) \; . \qquad (16)$$

For analysis purposes, we now define an additional SDMC $Q' : \mathcal{X} \to \mathcal{Z}$. Note that Q and Q' share the same output alphabet. The channel Q' is defined similarly, but with an "index shift". That is, define $\lambda_0 = 1$. Then,

$$Q'(z|1) = \begin{cases} \frac{\lambda_{i-1}}{M(\lambda_{i-1}+1)} & \text{if } z = z_i \; , \\ \frac{1}{M(\lambda_{i-1}+1)} & \text{if } z = \bar{z}_i \; , \end{cases} \qquad (17)$$

and

$$Q'(z_i|-1) = Q'(\bar{z}_i|1) \; , \quad Q'(\bar{z}_i|-1) = Q'(z_i|1) \; . \qquad (18)$$

Due to space constraints, we omit the proof of the following lemma.

Lemma 6. *The above channels satisfy $Q' \preceq W \preceq Q$.*

We are now ready to state our main result in this section, relating $\Psi(Q)$ to $I(W)$.

Theorem 7. *Let $W : \mathcal{X} \to \mathcal{Y}$ be a continuous channel as defined above. For a given integer M, let $Q : \mathcal{X} \to \mathcal{Z}$ be the upgraded channel described previously. Then, $|\mathcal{Z}| = 2M$ and*

$$\Psi(Q) - I(W) \leqslant \frac{1}{M} \; .$$

Proof: The claim $|\mathcal{Z}| = 2M$ is simply a restatement of (14). Next, note that for all $z \in \mathcal{Z}$

$$Q(z) = \frac{Q(z|1) + Q(z|-1)}{2} = \frac{1}{2M} \; .$$

Namely, a uniform input to Q results in a uniform output, and thus,

$$\Psi(Q) = I(Q) \; . \qquad (19)$$

Hence, we are to prove that

$$I(Q) - I(W) \leqslant \frac{1}{M} \; .$$

By Lemma 6 we have that

$$I(Q') \leqslant I(W) \leqslant I(Q) \; .$$

Thus, it suffices to prove that

$$I(Q) - I(Q') = \frac{1}{M} \; . \qquad (20)$$

To this end, define the function C as follows. For $0 \leqslant \lambda < \infty$

$$C[\lambda] = 1 - \frac{\lambda}{\lambda+1} \log_2 \left(1 + \frac{1}{\lambda}\right) - \frac{1}{\lambda+1} \log_2 (\lambda+1) \; ,$$

and (for continuity) we define $C[\infty] = 1$. Now, a short calculation shows that for any SDMC Q with output alphabet $\mathcal{Z} = \{z_1, \bar{z}_1, z_2, \bar{z}_2, \ldots, z_M, \bar{z}_M\}$ and input alphabet $\mathcal{X} = \{-1, 1\}$ we have that

$$I(Q) = \sum_{i=1}^{M} Q(Z) C[Q(z|1)/Q(z|-1)] \; . \qquad (21)$$

Specializing (21) to our Q gives

$$I(Q) = \frac{1}{M} \sum_{i=1}^{M} C[\lambda_i] \; ,$$

while specialzing (21) to Q' gives

$$I(Q') = \frac{1}{M} \sum_{i=1}^{M} C[\lambda_{i-1}] \; ,$$

Thus,

$$I(Q) - I(Q') = \frac{1}{M} \left(C[\lambda_M] - C[\lambda_0]\right) = \frac{1}{M} \; .$$

■

References

[1] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inform. Theory*, vol. 55, no. 7, pp. 3051-3073, July 2009.

[2] M. Bellare and S. Tessaro, "Polynomial-time semantically-secure encryption achieving the secrecy capacity," arXiv:1201.3160, September 2011.

[3] M. Bellare, S. Tessaro, and A. Vardy, "A cryptographic treatment of the wiretap channel," arXiv:1201.2205, April 2010.

[4] M. Bellare, S. Tessaro, and A. Vardy, "Semantic security for the wiretap channel," in *Proc. 32-nd Annual Cryptology Conference* (CRYPTO), Santa Barbara, CA., *Lect. Notes Computer Science*, vol. 7417, pp. 294–311, August 2012.

[5] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd ed., Cambridge, MA: The MIT Press, 2001.

[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., New York: John Wiley, 2006.

[7] R. G. Gallager, *Information Theory and Reliable Communications*. New York: John Wiley, 1968.

[8] S. Goldwasser and S. Micali, "Probabilistic encryption," *J. Computer and System Sciences*, vol. 28, pp. 270–299, 1984.

[9] H. Mahdavifar and A. Vardy, "Achieving the secrecy capacity of wiretap channels using polar codes," *IEEE Trans. Inform. Theory,* vol. 57, no.10, pp. 6428–6443, October 2011.

[10] E. Şaşoğlu and A. Vardy, "A new polar coding scheme for strong security on wiretap channels," submitted for presentation at the *IEEE Int. Symp. Inform. Theory*, Istanbul, Turkey, July 2013.

[11] I. Tal and A. Vardy, "How to construct polar codes," *submitted to IEEE Trans. Inform. Theory*, available as `arXiv:1105.6164v2`, May 2011.

[12] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54(8), pp. 1355–1387, 1975.