

The Secrecy Capacity of the Wiretap Channel with Rate-limited Feedback

Ehsan Ardetsanizadeh, Massimo Franceschetti, Tara Javidi, and Young-Han Kim

Abstract—This paper studies the problem of secure communication over a wiretap channel $p(y, z|x)$ with a secure feedback link of rate R_f , where X is the channel input, and Y and Z are channel outputs observed by the legitimate receiver and the eavesdropper, respectively. It is shown that the secrecy capacity, the maximum data rate of reliable communication while the intended message is not revealed to the eavesdropper, is upper bounded as

$$C_s(R_f) \leq \max_{p(x)} \min\{I(X; Y), I(X; Y|Z) + R_f\}.$$

The proof of the bound crucially depends on a recursive argument which is used to obtain the single-letter characterization. This upper bound is shown to be tight for the class of physically degraded wiretap channels. A capacity-achieving coding scheme is presented for this case, in which the receiver securely feeds back fresh randomness with rate R_f , generated *independent* of the received channel output symbols. The transmitter then uses this shared randomness as a secret key on top of Wyner's coding scheme for wiretap channels without feedback. Hence, when a feedback link is available, the receiver should allocate all resources to convey a new key rather than sending back the channel output.

I. INTRODUCTION

In his pioneering work [1] that opened up the era of modern cryptography, Shannon modeled a secrecy system as a communication system consisting of a legitimate transmitter (Alice), a legitimate receiver (Bob), and an eavesdropper (Eve), in which Alice wishes to transmit a message M to Bob secret from Eve. If Eve has complete access to what Bob receives, Shannon showed that in order to achieve *perfect secrecy*, a secret key K of entropy $H(K) \geq H(M)$ has to be shared between Alice and Bob. This fundamental yet strongly negative result has been extended—and in a sense overcome—in many directions. In the direction of mathematical communication theory, Wyner [2] introduced the degraded wiretap channel, in which Bob receives the message through a discrete memoryless channel (DMC) $p(y|x)$, and Eve has access to what Bob receives through an additional discrete memoryless channel $p(z|y)$ such that $p(y, z|x) = p(y|x)p(z|y)$, as depicted in Figure 1. By relaxing the secrecy requirement mildly while exploiting the better quality of the Alice–Bob channel $p(y|x)$ than that of the Alice–Eve channel $p(z|x)$, Wyner showed that information can be transmitted securely at a positive rate, and

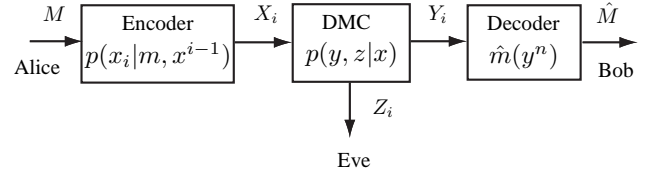


Fig. 1. The wiretap channel.

characterized the secrecy capacity C_s , the supremum of all achievable rates of secure communication, as

$$C_s = \max_{p(x)} I(X; Y|Z) = \max_{p(x)} (I(X; Y) - I(X; Z)). \quad (1)$$

This result was later extended by Csiszár and Körner [3] to general broadcast channels with confidential messages. In particular, they showed that the secrecy capacity of the general (not necessarily degraded) wiretap channel $p(y, z|x)$ is

$$C_s = \max_{p(u, x)} (I(U; Y) - I(U; Z)). \quad (2)$$

Furthermore, it was shown that if the channel from Alice to Bob is *more capable* [4] than the channel from Alice to Eve, that is, if $I(X; Y) \geq I(X; Z)$ for all $p(x)$, then the secrecy capacity is simplified to

$$C_s = \max_{p(x)} (I(X; Y) - I(X; Z)). \quad (3)$$

All the scenarios described above deal with one-way communications between Alice and Bob. However, many common communications arise over inherently two-way channels, such as telephone systems, digital subscriber lines (DSL), cellular networks, satellite communications, and the Internet. Hence, it is natural to ask how possible interactions between Alice and Bob can increase the secrecy of their communication.

As a canonical model to study this question, this paper extends the wiretap channel model by introducing a secure feedback link of rate R_f from Bob to Alice as depicted in Figure 2. The secure feedback link can be viewed as a primitive form of the backward channel from Bob to Alice with secrecy capacity R_f , independent of the forward channel. Thus this model can provide insights into the value of two-way interactions in secure communication.

There are several concrete scenarios in which this model is applicable. For instance, consider the communication between a satellite (Alice) and a base station (Bob) on the ground. The satellite broadcasts its signal to the ground, so any (unintended) station can receive it. On the other hand, the base station can beamform some data back to the satellite securely,

The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Toronto, Canada, July 2008.

The authors are with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA, 92093-0407, USA (e-mail: {eardesta, mfranceschetti, tjavidi, yhk}@ucsd.edu).

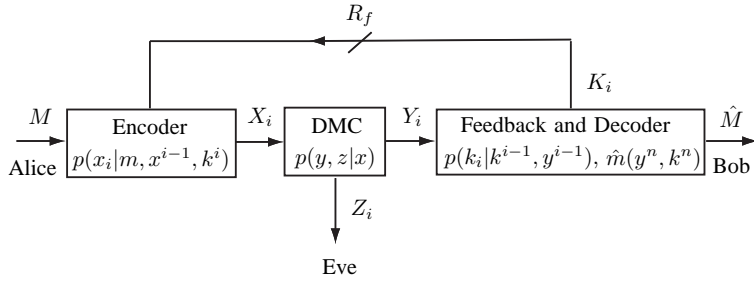


Fig. 2. The wiretap channel with secure rate-limited feedback.

which can be used to enhance the secret data rate sent from the satellite to the base station.

The main purpose of this paper is to investigate the secrecy capacity $C_s(R_f)$ as a function of the secure feedback rate R_f . In Theorem 1, we show the following upper bound for a general wiretap channel $p(y, z|x)$ with secure rate-limited feedback:

$$C_s(R_f) \leq \max_{p(x)} \min\{I(X; Y), I(X; Y|Z) + R_f\}. \quad (4)$$

Due to the dependencies introduced by the feedback, proving the upper bound (4) requires a less standard treatment. We use a *recursive* argument, which helps to track the causal dependencies step by step, and obtain the desired single-letter characterization. Exploiting the recursive structure to find the single-letter characterization might be a powerful tool for similar proofs.

For the case of a physically degraded wiretap channel $p(y, z|x) = p(y|x)p(z|y)$, in which Eve receives a degraded version of what Bob receives, we show that the upper bound (4) is tight, establishing the secrecy capacity as

$$C_s(R_f) = \max_{p(x)} \min\{I(X; Y), I(X; Y|Z) + R_f\}. \quad (5)$$

Interestingly, we show that in order to achieve the secrecy capacity for the physically degraded wiretap channel Bob can simply ignore what he receives and sends “fresh” randomness. This fresh randomness plays the role of a secret key, which bridges Shannon’s original result with Wyner’s wiretap model. Accordingly, we modify Wyner’s original coding scheme to allow the use of a shared key (sent from Bob to Alice via rate-limited feedback). It should be noted that this modification has been already proposed by Yamamoto [7] and Merhav [8], who characterized the secrecy capacity of wiretap channels with a shared key (which is already given prior to the communication) and also considered additional effects of having distortion or side information.

In a closely related work, Ahlswede and Cai [6] studied wiretap channels with secure *output* feedback, in which the channel output symbols received by Bob are fed back secretly to Alice. They showed that the secrecy capacity of the physically degraded wiretap channel with secure output feedback is

$$\max_{p(x)} \min\{I(X; Y), I(X; Y|Z) + H(Y|X, Z)\}, \quad (6)$$

which is in general larger than the nonfeedback secrecy capacity (1). At a first glance, it might seem contradictory that the optimal receiver should ignore the channel outputs completely when feedback is rate-limited (in the current paper) while the output feedback (in the Ahlswede–Cai setup) boosts the secrecy capacity as in (6). A closer look, however, reveals that the Ahlswede–Cai coding scheme essentially extracts fresh randomness in the feedback output symbols hidden from Eve and uses that randomness as a key. Hence, our result shows explicitly that when Bob has a means of interacting with Alice, he should allocate all resources to convey a key rather than sending back the channel output.

Recently, additional studies have been conducted on characterizing the secrecy capacity of various two-way communication systems. Lai, El Gamal, and Poor [9] studied the case of the modulo-additive DMC, where Eve receives the modulo-sum of the source signal, the feedback signal, and the noise. They showed that if Bob jams Eve completely, then Alice can send messages securely at the capacity of the channel to Bob. Tekin and Yener [10] presented an achievable rate region for the two-way Gaussian wiretap channel. Similar to [9], the model presented in [10] assumes that Eve receives the sum of signals from both transmitters corrupted by an additive Gaussian noise. They showed that due to the multiple access nature of Eve’s channel, each transmitter can simultaneously help to hide the other user’s message from Eve and send some data secretly to the other user. In both studies, the additive nature of Eve’s channel gives the opportunity for jamming, in addition to possible backward information transfer. In comparison, our model decouples the forward and backward communication channels, eliminating the possible use of jamming, and focuses on the “inherent” value of the backward communication link. It also seems that having independent forward and backward communication links fits better the current practice of two-way communications over orthogonal media such as different frequency bands or time slots.

The rest of the paper is organized as follows. First, we give a formal statement of our result in Section II. Then, we show the upper bound on the secrecy capacity and the coding scheme in Sections III and IV, respectively. Section V concludes the paper.

II. PROBLEM SETUP AND THE MAIN RESULT

We consider the communication problem depicted in Figure 2. Here Alice communicates a message index $M \in [2^{nR}] := \{1, 2, \dots, 2^{nR}\}$ over a wiretap channel $p(y, z|x)$, where the channel input $X_i \in \mathcal{X}$ at time i is received as $Y_i \in \mathcal{Y}$ and $Z_i \in \mathcal{Z}$, respectively, by the legitimate receiver Bob and the eavesdropper Eve. Alice wishes to communicate the message M to Bob reliably over their channel $p(y|x)$ while keeping it secret from Eve. To enhance the secrecy of the communication, Bob can communicate back symbols $K_i \in \mathcal{K}_i$, $i = 1, 2, \dots, n$, over a feedback link of rate R_f secret from Eve. The feedback symbol K_i at time i can depend causally on previous channel outputs $Y^{i-1} := (Y_1, \dots, Y_{i-1})$ and previous feedback symbols $K^{i-1} := (K_1, \dots, K_{i-1})$. We assume that the channel alphabets $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$, and the feedback alphabets $\mathcal{K}_1, \dots, \mathcal{K}_n$ are finite, and Eve has complete knowledge about them as well as the coding scheme used by Alice and Bob. The wiretap channel $p(y, z|x)$ is memoryless, i.e.,

$$p(y_i, z_i|x^i, y^{i-1}, z^{i-1}) = p(y_i, z_i|x_i)$$

for $i = 1, 2, \dots, n$.

More formally, we define a $(2^{nR}, 2^{nR_f}, n)$ code as

- 1) feedback alphabets $\mathcal{K}_1, \dots, \mathcal{K}_n$ such that their cardinalities satisfy

$$\frac{1}{n} \sum_{i=1}^n \log(|\mathcal{K}_i|) \leq R_f, \quad (7)$$

- 2) stochastic encoding maps consisting of conditional probability distributions $p(x_i|m, x^{i-1}, k^i)$, $i = 1, 2, \dots, n$, defined for each $x_i \in \mathcal{X}$, $k^i \in \mathcal{K}^i := \mathcal{K}_1 \times \dots \times \mathcal{K}_i$, $x^{i-1} \in \mathcal{X}^{i-1}$, and $m \in [2^{nR}]$ such that for each i, m, x^{i-1}, k^i , $\sum_{x_i} p(x_i|m, x^{i-1}, k^i) = 1$ (in other words, $p(x_i|m, x^{i-1}, k^i)$ denotes the probability that the message m , the previous sent symbols x^{i-1} and the previously received feedback symbols k^i are mapped to the channel input x_i at time i),
- 3) stochastic feedback maps consisting of conditional probability distributions $p(k_i|y^{i-1}, k^{i-1})$ (by convention, $K_1 \sim p(k_1)$ independent of M), and
- 4) a decoding map $\hat{m}: \mathcal{Y}^n \times \mathcal{K}^n \rightarrow [2^{nR}]$ resulting in the decoded message

$$\hat{M} = \hat{m}(Y^n, K^n). \quad (8)$$

We assume throughout that the message M is a random variable uniformly distributed over $[2^{nR}]$. Given a $(2^{nR}, 2^{nR_f}, n)$ code, we define the probability of error $P_e^{(n)}$ as

$$P_e^{(n)} := \Pr\{\hat{M} \neq M\},$$

and the secrecy measure $L^{(n)}$ as

$$L^{(n)} := \frac{1}{n} I(M; Z^n).$$

Definition 1: A secrecy rate R is achievable if there exists a sequence of $(2^{nR}, 2^{nR_f}, n)$ codes such that as $n \rightarrow \infty$,

$$P_e^{(n)} \rightarrow 0, \quad (9)$$

$$L^{(n)} \rightarrow 0. \quad (10)$$

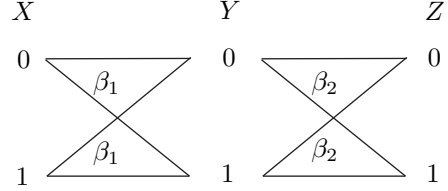


Fig. 3. The physically degraded binary symmetric wiretap channel.

Note that $L^{(n)} = R(1 - \Delta^{(n)})$, where

$$\Delta^{(n)} = \frac{H(M|Z^n)}{H(M)} = \frac{H(M|Z^n)}{nR},$$

is the *equivocation* as was defined originally by Wyner, and the condition $L^{(n)} \rightarrow 0$ in Definition 1 is equivalent to the condition $\Delta^{(n)} \rightarrow 1$, which was used by Wyner as the requirement for secure communication. The secrecy capacity $C_s(R_f)$ at feedback rate R_f is the supremum of all achievable secrecy rates. We are now ready to state our main results.

Theorem 1: The secrecy capacity $C_s(R_f)$ of the wiretap channel with rate-limited feedback R_f is upper bounded as

$$C_s(R_f) \leq \max_{p(x)} \min\{I(X; Y), I(X; Y|Z) + R_f\}.$$

The proof is given in Section III.

Theorem 2: The secrecy capacity of the physically degraded wiretap channel $p(y, z|x) = p(y|x)p(z|y)$ with rate-limited feedback R_f is

$$C_s(R_f) = \max_{p(x)} \min\{I(X; Y), I(X; Y|Z) + R_f\}. \quad (11)$$

The converse follows immediately from Theorem 1. A capacity-achieving coding scheme is presented in Section IV, in which Bob sends back pure randomness securely at rate R_f , and Alice uses that shared randomness to increase the secrecy rate.

Example 1: Consider the degraded wiretap channel shown in Figure 3, which is a cascade of two binary symmetric channels, $\text{BSC}(\beta_1)$ and $\text{BSC}(\beta_2)$. By symmetry, the distribution $\Pr(X = 0) = \Pr(X = 1) = \frac{1}{2}$ achieves the maximization in $C_s(R_f)$, and with this distribution we have

$$I(X; Y) = 1 - h(\beta_1)$$

$$I(X; Y|Z) = I(X; Y) - I(X; Z) = h(\beta_1 * \beta_2) - h(\beta_1),$$

where $\beta_1 * \beta_2 = \beta_1(1 - \beta_2) + (1 - \beta_1)\beta_2$.

From Theorem 2, we have

$$C_s(R_f) = \min\{1 - h(\beta_1), h(\beta_1 * \beta_2) - h(\beta_1) + R_f\}.$$

Figure 7 shows the plot of $C_s(R_f)$, which starts from $C_s(0) = h(\beta_1 * \beta_2) - h(\beta_1)$ and increases linearly with R_f until it gets saturated at $C = 1 - h(\beta_1)$, for feedback rate $R_f \geq 1 - h(\beta_1 * \beta_2)$.

Example 2: In this example we look at the physically degraded Gaussian wiretap channel shown in Figure 5. Here E_1 and E_2 are assumed to be independent from each other, i.i.d. over time, and distributed as $E_1 \sim \mathcal{N}(0, N_1)$ and

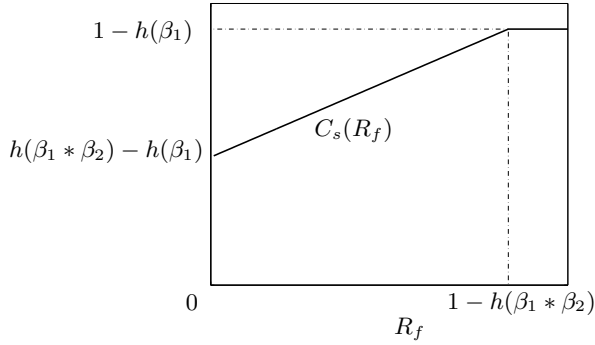


Fig. 4. Plot of $C_s(R_f)$ for Example 1.

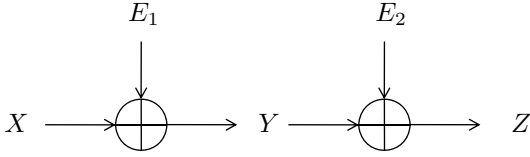


Fig. 5. The physically degraded Gaussian wiretap channel.

$E_2 \sim \mathcal{N}(0, N_2)$, where $\mathcal{N}(0, N)$ is a Gaussian distribution with zero mean and variance N .

Let P be the input power constraint. Then we know [11]

$$C = \max_{E_{X^2} \leq P} I(X; Y) = \frac{1}{2} \log\left(1 + \frac{P}{N_1}\right), \quad (12)$$

where the maximum is attained by $X \sim \mathcal{N}(0, P)$. For the secrecy capacity without feedback we know [5] that

$$\begin{aligned} C_s(0) &= \max_{E_{X^2} \leq P} I(X; Y|Z) \\ &= \frac{1}{2} \log\left(1 + \frac{P}{N_1}\right) - \frac{1}{2} \log\left(1 + \frac{P}{N_1 + N_2}\right), \end{aligned} \quad (13)$$

where the maximum is attained again with $X \sim \mathcal{N}(0, P)$. We will not provide the argument; however, it is straightforward to show that Theorem 2 can be modified for additive Gaussian noise channels to give

$$C_s(R_f) = \max_{E_{X^2} \leq P} \min\{I(X; Y), I(X; Y|Z) + R_f\}. \quad (14)$$

Since the maximizations in (12) and (13) are achieved by the same distribution, it can be verified that the maximization in (14) is also achieved by $X \sim \mathcal{N}(0, P)$ and the secrecy capacity with rate-limited feedback is

$$C_s(R_f) = \min\{C, C_s(0) + R_f\}. \quad (15)$$

Similar to Figure 4 in the previous example, $C_s(R_f)$ starts from $C_s(0) = \frac{1}{2} \log\left(1 + \frac{P}{N_1}\right) - \frac{1}{2} \log\left(1 + \frac{P}{N_1 + N_2}\right)$ and increases linearly with R_f until it gets saturated at $C = \frac{1}{2} \log\left(1 + \frac{P}{N_1}\right)$, for feedback rate $R_f \geq \frac{1}{2} \log\left(1 + \frac{P}{N_1 + N_2}\right)$.

As we saw in the previous examples, when the same input distribution maximizes $I(X; Y)$ and $I(X; Y|Z)$, the maximization and the minimization in $C_s(R_f)$ can be exchanged. Therefore,

$$C_s(R_f) = \min\{C, C_s(0) + R_f\},$$

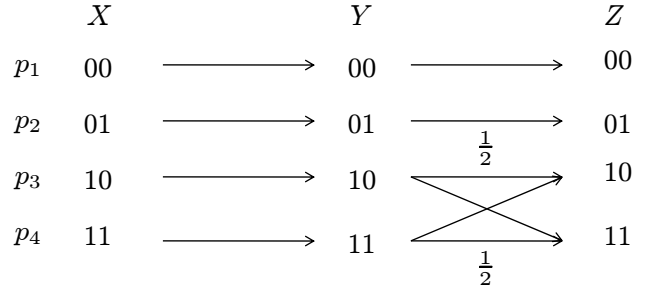


Fig. 6. An example of a degraded wiretap channel whose secrecy capacity is sublinear in the feedback rate.

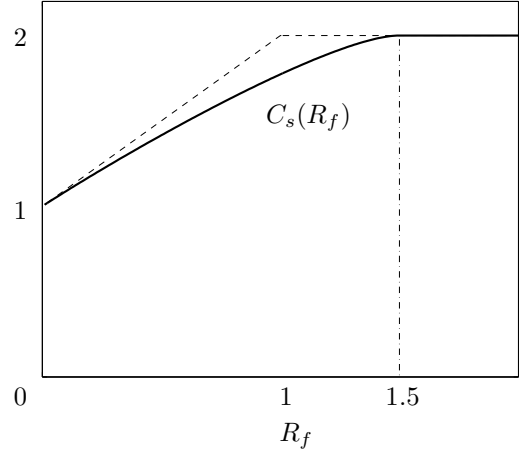


Fig. 7. Plot of $C_s(R_f)$ for Example 3. The dashed line shows $\min\{C, C_s(0) + R_f\}$.

and one bit secure feedback is worth one bit in the secrecy capacity until we get limited by the capacity of the channel between Alice and Bob. However, this is not always true. In fact, $C_s(R_f)$ could be strictly sublinear in R_f ; namely,

$$C_s(R_f) < \min\{C, C_s(0) + R_f\},$$

as shown in the following example.

Example 3: Consider the degraded wiretap channel shown in Figure 6. By symmetry, the distribution $p_1 = p_2 = \frac{p}{2}$ and $p_3 = p_4 = \frac{1-p}{2}$ achieves the maximization in (11). It is easy to verify that with this input distribution we have

$$I(X; Y) = h(p) + 1 \quad (16)$$

$$I(X; Y|Z) = p \quad (17)$$

where $h(p) = -p \log p - (1-p) \log(1-p)$ is the binary entropy function.

It follows that

$$C_s(R_f) = \max_p \min\{h(p) + 1, p + R_f\}. \quad (18)$$

Figure 7 shows the plot of $C_s(R_f)$, which increases sublinearly with R_f until it gets saturated at $C = 2$, for feedback rate $R_f \geq 1.5$.

III. PROOF OF THE UPPER BOUND

In this section we show that if the secrecy rate R is achievable, then R must satisfy

$$R \leq \max_{p(x)} \min\{I(X; Y), I(X; Y|Z) + R_f\}. \quad (19)$$

To show (19) we prove the following two upper bounds for any achievable secrecy rate R .

$$R \leq \frac{1}{n} \sum_{i=1}^n I(X_i; Y_i) + \epsilon_n \quad (20)$$

$$R \leq R_f + \frac{1}{n} \sum_{i=1}^n I(X_i; Y_i|Z_i) + \delta_n, \quad (21)$$

where $\epsilon_n, \delta_n \rightarrow 0$ as $n \rightarrow \infty$. Then we use the usual technique of introducing a time sharing random variable [11], and concavity of mutual information in $p(x)$ to obtain (19).

First, (20) follows easily from Fano's inequality as in the standard converse proof of the channel coding theorem [11, Theorem 7.7.1].

We now prove (21) using Fano's inequality, the secrecy constraint (10), and the feedback rate-limit constraint (7). A recursive argument (Lemma 3) is then used to obtain the single-letter characterization.

By Fano's inequality, we have

$$H(M|\hat{M}) \leq 1 + P_e^{(n)} nR =: n\epsilon_n.$$

By the assumption that $P_e^{(n)} \rightarrow 0$, we have $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$. From (8) and the data processing inequality, we have

$$H(M|K^n, Y^n) \leq H(M|\hat{M}) \leq n\epsilon_n.$$

By the assumption that $L^{(n)} \rightarrow 0$, we have

$$I(M; Z^n) = n\gamma_n, \quad (22)$$

where $\gamma_n \rightarrow 0$ as $n \rightarrow \infty$. It then follows that

$$\begin{aligned} nR &= H(M) \\ &= H(M|Z^n) + I(M; Z^n) \\ &= H(M|Z^n) + n\gamma_n \end{aligned} \quad (23)$$

$$\begin{aligned} &= I(M; Y^n, K^n|Z^n) + H(M|Y^n, Z^n, K^n) + n\gamma_n \\ &\leq I(M; Y^n, K^n|Z^n) + n\epsilon_n + n\gamma_n \end{aligned} \quad (24)$$

$$= I(M; K^n|Z^n) + I(M; Y^n|K^n, Z^n) + n\delta_n \quad (25)$$

$$\leq H(K^n|Z^n) + I(M, X^n; Y^n|K^n, Z^n) + n\delta_n, \quad (26)$$

where (23) follows from (22), (24) follows from Fano's inequality, (25) follows by defining $\delta_n = \epsilon_n + \gamma_n$.

The following lemma provides a recursive expression, which is crucial to single-letterize (26).

Lemma 3: For each $j = 1, 2, \dots, n$, we have

$$\begin{aligned} &H(K^j|Z^j) + I(M, X^j; Y^j|K^j, Z^j) \\ &\leq H(K^{j-1}|Z^{j-1}) + I(M, X^{j-1}; Y^{j-1}|K^{j-1}, Z^{j-1}) \\ &\quad + H(K_j|M, X^{j-1} + K^{j-1}, Z^{j-1}) + I(X_j; Y_j|Z_j). \end{aligned}$$

Proof: We have the following chain of inequalities:

$$\begin{aligned} &H(K^j|Z^j) + I(M, X^j; Y^j|K^j, Z^j) \\ &= H(K^j|Z^j) + I(M, X^j; Y^{j-1}|K^j, Z^j) \end{aligned}$$

$$\begin{aligned} &+ I(M, X^j; Y_j|Y^{j-1}, K^j, Z^j) \\ &\leq H(K^j|Z^j) + I(M, X^j; Y^{j-1}|K^j, Z^j) \\ &\quad + I(M, Y^{j-1}, K^j, Z^{j-1}; X^j; Y_j|Z_j) \\ &= H(K^j|Z^j) + I(M, X^j; Y^{j-1}|K^j, Z^j) \\ &\quad + I(X_j; Y_j|Z_j) \end{aligned} \quad (27)$$

$$\begin{aligned} &\leq H(K^j|Z^j) + I(M, X^j, Z_j; Y^{j-1}|K^j, Z^{j-1}) \\ &\quad + I(X_j; Y_j|Z_j) \\ &= H(K^j|Z^j) + I(M, X^j; Y^{j-1}|K^j, Z^{j-1}) \\ &\quad + I(X_j; Y_j|Z_j) \end{aligned} \quad (28)$$

$$\begin{aligned} &= H(K^j|Z^j) + I(M, X^{j-1}; Y^{j-1}|K^j, Z^{j-1}) \\ &\quad + I(X_j; Y^{j-1}|M, X^{j-1}, K^j, Z^{j-1}) \\ &\quad + I(X_j; Y_j|Z_j) \\ &= H(K^j|Z^j) + I(M, X^{j-1}; Y^{j-1}|K^j, Z^{j-1}) \\ &\quad + I(X_j; Y_j|Z_j) \end{aligned} \quad (29)$$

$$\begin{aligned} &= H(K^j|Z^j) + I(M, X^{j-1}, K_j; Y^{j-1}|K^{j-1}, Z^{j-1}) \\ &\quad - I(K_j; Y^{j-1}|K^{j-1}, Z^{j-1}) + I(X_j; Y_j|Z_j) \\ &= H(K^j|Z^j) + I(M, X^{j-1}; Y^{j-1}|K^{j-1}, Z^{j-1}) \\ &\quad + I(K_j; Y^{j-1}|M, X^{j-1}, K^{j-1}, Z^{j-1}) \\ &\quad - I(K_j; Y^{j-1}|K^{j-1}, Z^{j-1}) + I(X_j; Y_j|Z_j) \end{aligned}$$

$$\begin{aligned} &= H(K^{j-1}|Z^j) + H(K_j|K^{j-1}, Z^j) \\ &\quad + I(M, X^{j-1}; Y^{j-1}|K^{j-1}, Z^{j-1}) \\ &\quad + I(K_j; Y^{j-1}|M, X^{j-1}, K^{j-1}, Z^{j-1}) \\ &\quad + H(K_j|Y^{j-1}, K^{j-1}, Z^{j-1}) \\ &\quad - H(K_j|K^{j-1}, Z^{j-1}) + I(X_j; Y_j|Z_j) \\ &\leq H(K^{j-1}|Z^j) + I(M, X^{j-1}; Y^{j-1}|K^{j-1}, Z^{j-1}) \\ &\quad + I(K_j; Y^{j-1}|M, X^{j-1}, K^{j-1}, Z^{j-1}) \\ &\quad + H(K_j|Y^{j-1}, K^{j-1}, Z^{j-1}) + I(X_j; Y_j|Z_j) \end{aligned} \quad (30)$$

$$\begin{aligned} &= H(K^{j-1}|Z^j) + I(M, X^{j-1}; Y^{j-1}|K^{j-1}, Z^{j-1}) \\ &\quad + H(K_j|M, X^{j-1}, K^{j-1}, Z^{j-1}) \\ &\quad - H(K_j|Y^{j-1}, M, X^{j-1}, K^{j-1}, Z^{j-1}) \\ &\quad + H(K_j|Y^{j-1}, K^{j-1}, Z^{j-1}) \\ &\quad + I(X_j; Y_j|Z_j) \\ &= H(K^{j-1}|Z^j) + I(M, X^{j-1}; Y^{j-1}|K^{j-1}, Z^{j-1}) \\ &\quad + H(K_j|M, X^{j-1}, K^{j-1}, Z^{j-1}) + I(X_j; Y_j|Z_j) \end{aligned} \quad (31)$$

$$\begin{aligned} &\leq H(K^{j-1}|Z^{j-1}) + I(M, X^{j-1}; Y^{j-1}|K^{j-1}, Z^{j-1}) \\ &\quad + H(K_j|M, X^{j-1}, K^{j-1}, Z^{j-1}) + I(X_j; Y_j|Z_j), \end{aligned} \quad (32)$$

where

- (27) holds because the channel is memoryless and therefore $Y_j \rightarrow (X_j, Z_j) \rightarrow (M, X^{j-1}, K^j, Y^{j-1}, Z^{j-1})$ form a Markov chain,
- (28) holds because $Z_j \rightarrow (M, X^j, K^j, Z^{j-1}) \rightarrow Y^{j-1}$ form a Markov chain,
- (29) holds because $Y^{j-1} \rightarrow (M, X^{j-1}, K^j, Z^{j-1}) \rightarrow X_j$ form a Markov chain,
- (30) and (32) follow from $H(K_j|K^{j-1}, Z^j) \leq H(K_j|K^{j-1}, Z^{j-1})$ and $H(K_j|Z^j) \leq H(K_j|Z^{j-1})$

respectively,

- and (31) holds because of the following Markov chain $(M, X^{j-1}, Z^{j-1}) \rightarrow (Y^{j-1}, K^{j-1}) \rightarrow K_j$. ■

Starting from (26), we apply Lemma 3 recursively to find the single-letter characterization as follows:

$$\begin{aligned}
nR &\leq H(K^n|Z^n) + I(M, X^n; Y^n|K^n, Z^n) + n\delta_n \\
&\leq H(K^{n-1}|Z^{n-1}) + I(M, X^{n-1}; Y^{n-1}|K^{n-1}, Z^{n-1}) \\
&\quad + I(X_n; Y_n|Z_n) + H(K_n) + n\delta_n \\
&\leq H(K^{n-2}|Z^{n-2}) + I(M, X^{n-2}; Y^{n-2}|K^{n-2}, Z^{n-2}) \\
&\quad + I(X_{n-1}; Y_{n-1}|Z_{n-1}) + H(K_{n-1}) \\
&\quad + I(X_n; Y_n|Z_n) + H(K_n) + n\delta_n \\
&\quad \vdots \\
&\leq \sum_{i=1}^n I(X_i; Y_i|Z_i) + \sum_{i=1}^n H(K_i) + n\delta_n. \tag{33}
\end{aligned}$$

Dividing by n and applying the feedback rate-limit constraint (7) we obtain (21) as follows:

$$\begin{aligned}
R &\leq \frac{1}{n} \sum_{i=1}^n I(X_i; Y_i|Z_i) + \frac{1}{n} \sum_{i=1}^n H(K_i) + \delta_n \\
&\leq \frac{1}{n} \sum_{i=1}^n I(X_i; Y_i|Z_i) + R_f + \delta_n.
\end{aligned}$$

To complete the proof, let Q be a time-sharing random variable distributed uniformly over $\{1, 2, \dots, n\}$ and independent of X^n, Y^n, Z^n . Then (21) can be written as

$$\begin{aligned}
R &\leq R_f + \frac{1}{n} \sum_{i=1}^n I(X_i; Y_i|Z_i) + \delta_n \\
&= R_f + \frac{1}{n} \sum_{i=1}^n I(X_i; Y_i|Z_i, Q=i) + \delta_n \\
&= R_f + I(X_Q; Y_Q|Z_Q, Q) + \delta_n \\
&= R_f + I(X; Y|Z, Q) + \delta_n,
\end{aligned}$$

where $X \triangleq X_Q, Y \triangleq Y_Q, Z \triangleq Z_Q$. Now letting $n \rightarrow \infty, \delta_n \rightarrow 0$ and hence

$$R \leq R_f + I(X; Y|Z, Q). \tag{34}$$

Similarly, (20) can be written as

$$R \leq I(X; Y|Q) + \epsilon_n,$$

where as $n \rightarrow \infty, \epsilon_n \rightarrow 0$ and we have

$$R \leq I(X; Y|Q). \tag{35}$$

Note that $\Pr(Y_Q = y, Z_Q = z|X_Q = x)$ is consistent with the given wiretap channel $p(y, z|x)$ and is independent of Q . Since $Q \rightarrow X \rightarrow Y \rightarrow Z$ form a Markov chain, it follows from (34),

$$\begin{aligned}
R &\leq R_f + I(X; Y|Z, Q) \\
&\leq R_f + I(X, Q; Y|Z) \\
&= R_f + I(X; Y|Z). \tag{36}
\end{aligned}$$

Similarly from (35),

$$R \leq I(X; Y|Q) \leq I(X, Q; Y) = I(X; Y). \tag{37}$$

Combining (36) and (37), we have

$$R \leq \min[I(X; Y), I(X; Y|Z) + R_f]$$

for some (X, Y, Z) consistent with the given channel $p(y, z|x)$. Therefore, we conclude that

$$R \leq \max_{p(x)} \min[I(X; Y), I(X; Y|Z) + R_f],$$

which completes the proof of Theorem 1.

IV. A CAPACITY-ACHIEVING SCHEME FOR DEGRADED CHANNELS

In this section we present a coding scheme for the degraded wiretap channel with rate-limited feedback that achieves any secrecy rate R satisfying

$$R < \max_{p(x)} \min [I(X; Y), I(X; Y|Z) + R_f]. \tag{38}$$

We assume Bob uses the feedback link only to send back a secret key of rate R_f at the first instance; i.e., send K_1 with $|\mathcal{K}_1| = 2^{nR_f}$, so that Alice and Bob have a shared key prior to their communication as shown in Figure 8. In the remainder, we will provide a coding scheme for the wiretap channel with shared key of rate R_f that achieves any secrecy rate R satisfying (38).

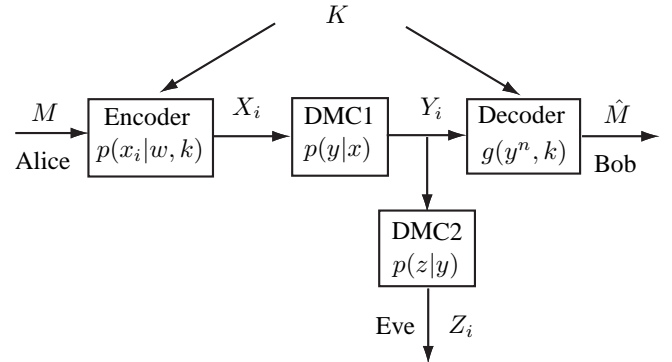


Fig. 8. The degraded wiretap channel with shared key.

Fix any distribution $p(x)$ and define

$$R' = I(X; Y|Z) = I(X; Y) - I(X; Z), \tag{39}$$

and

$$R'_f = R - R'. \tag{40}$$

Let $M = \{M_1, M_2\}$, where M_1 and M_2 are independent random variables uniformly distributed over $[2^{nR'}]$ and $[2^{nR'_f}]$, respectively. As shown below, the message M_1 will be transmitted securely using Wyner's original coding scheme while the security of M_2 will be guaranteed by using a key of rate R'_f . Note that from (38), (39) and (40) we have

$$R'_f < \min[I(X; Z), R_f].$$

The codebook is a collection of codewords $X^n \in \mathcal{X}^n$, from which the specific codeword $X^n(M_1, M_2, K)$ is picked randomly so as to confuse the eavesdropper. Therefore, there is no pre-defined codeword for a specific message.

Codebook generation. Pick $R_C \geq R$, such that $R_C = I(X; Y) - \epsilon$ for some $\epsilon > 0$. Such (R_C, ϵ) always exists since $R < I(X; Y)$. Generate a random codebook \mathcal{C} containing i.i.d. random codewords $X^n(\ell) \in \mathcal{X}^n$, $\ell \in [2^{nR_C}]$, each drawn according to $\Pr(X^n = x^n) = \prod_{i=1}^n p(x_i)$. Divide the codebook into $2^{nR'}$ disjoint *sub-codebooks*, each of which has $2^{n(R_C - R')}$ codewords. Label the sub-codebooks $\mathcal{C}_1, \dots, \mathcal{C}_{2^{nR'}}$. Now, divide each sub-codebook \mathcal{C}_i into $T = 2^{n(R_C - R' - R'_f)}$ sections $\mathcal{C}_{i1}, \dots, \mathcal{C}_{iT}$, each of which has $2^{nR'_f}$ codewords. Enumerate the codewords in each section from 1 to $2^{nR'_f}$, so the codewords in the j -th section of the i -th sub-codebook can be called as $X_{\mathcal{C}_{ij}}^n(1), \dots, X_{\mathcal{C}_{ij}}^n(2^{nR'_f})$ (see Figure 9).

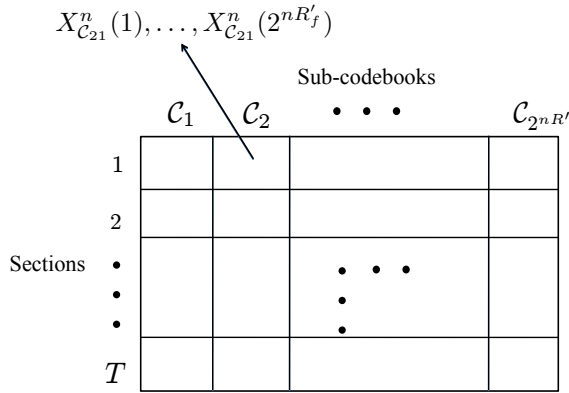


Fig. 9. Structure of the codebook.

Feedback. Let K be uniform over $[2^{nR'_f}]$. Bob sends $K_1 = K$ at time 1.

Encoding. We use K as a key shared between Alice and Bob. Generate a new variable $M'_2 = M_2 \oplus K \in [2^{nR'_f}]$, where \oplus is the modulo addition over the set $[2^{nR'_f}]$. Note that K and M_2 are uniformly distributed and independent, so M'_2 is uniformly distributed and independent of both K and M_2 .

We pick $X^n(M_1, M_2, K)$ as follows. According to M_1 we pick the corresponding sub-codebook among $2^{nR'}$ ones. In that sub-codebook we pick one of the sections uniformly randomly and in that section according to M'_2 we pick the corresponding codeword among the $2^{nR'_f}$ codewords in that section. We denote by $J \in [2^{n(R_C - R')}]$ the index of the picked codeword in the sub-codebook corresponding to M_1 .

Decoding. Decoder looks for a unique index $\hat{\ell} \in [2^{nR_C}]$ such that $(X^n(\hat{\ell}), Y^n) \in A_\epsilon^{(n)}$, where $A_\epsilon^{(n)}$ is the set of jointly typical (X^n, Y^n) sequences. If no such $\hat{\ell}$ exists or if there is more than one such, an error is declared. Having found $\hat{\ell}$, the decoder finds the reconstructed message (\hat{M}_1, \hat{M}_2) as follows. It chooses \hat{M}_1 as the index of the sub-codebook $X^n(\hat{\ell})$ belongs to. For \hat{M}_2 , the decoder first finds \hat{M}'_2 , the index of $X^n(\hat{\ell})$ in the section it belongs to, and then it finds $\hat{M}_2 = \hat{M}'_2 \ominus K$, where \ominus is the modulo subtraction over $[2^{nR'_f}]$.

Analysis of the error probability and the secrecy. We show

that there exists a codebook in the random collection of codebooks for which (9) and (10) are simultaneously satisfied; i.e., $P_e^{(n)} \rightarrow 0$ and $L^{(n)} = \frac{1}{n}I(M; Z^n) \rightarrow 0$.

Let $P_e^{(n)}(\mathcal{C}_0)$ and $L^{(n)}(\mathcal{C}_0)$ be the probability of error and the secrecy measure corresponding to a specific codebook \mathcal{C}_0 . Since $R_C < I(X; Y)$, by channel coding theorem [11, Theorem 7.7.1] we have

$$\mathbb{E}_{\mathcal{C}}[P_e^{(n)}(\mathcal{C})] \rightarrow 0 \text{ as } n \rightarrow \infty, \quad (41)$$

where the expectation is over all random codebooks. As shown in the following,

$$\mathbb{E}_{\mathcal{C}}[L^{(n)}(\mathcal{C})] \rightarrow 0 \text{ as } n \rightarrow \infty. \quad (42)$$

Combining (41) and (42), we have

$$\mathbb{E}_{\mathcal{C}}[P_e^{(n)}(\mathcal{C}) + L^{(n)}(\mathcal{C})] \rightarrow 0 \text{ as } n \rightarrow \infty. \quad (43)$$

Therefore, there exists at least one codebook for which conditions (9) and (10) are simultaneously satisfied. Now, it remains to show that (42) holds.

Let

$$L_1^{(n)}(\mathcal{C}_0) = \frac{1}{n}I(M_1; Z^n | \mathcal{C} = \mathcal{C}_0)$$

$$L^{(n)}(\mathcal{C}_0) = \frac{1}{n}I(M_1, M_2; Z^n | \mathcal{C} = \mathcal{C}_0).$$

For the rest of the paper all expectations are with respect to \mathcal{C} , so we omit the subscript \mathcal{C} . From the definition of the conditional mutual information we have

$$\mathbb{E}[L_1^{(n)}(\mathcal{C})] = \frac{1}{n}I(M_1; Z^n | \mathcal{C})$$

$$\mathbb{E}[L^{(n)}(\mathcal{C})] = \frac{1}{n}I(M_1, M_2; Z^n | \mathcal{C}). \quad (44)$$

With these definitions we have the following lemma.

Lemma 4: Suppose $\mathbb{E}[L_1^{(n)}(\mathcal{C})] \rightarrow 0$ as $n \rightarrow \infty$. Then $\mathbb{E}[L^{(n)}(\mathcal{C})] \rightarrow 0$ as $n \rightarrow \infty$.

Proof: Consider

$$\begin{aligned} \mathbb{E}[L^{(n)}(\mathcal{C})] &= \frac{1}{n}I(M_1, M_2; Z^n | \mathcal{C}) \\ &= \frac{1}{n}[I(M_1; Z^n | \mathcal{C}) + I(M_2; Z^n | \mathcal{C}, M_1)] \\ &= \frac{1}{n}I(M_1; Z^n | \mathcal{C}) \\ &= \mathbb{E}[L_1^{(n)}(\mathcal{C})], \end{aligned} \quad (45)$$

where (45) follows from the fact that M_2 is independent of (M_1, Z^n) for any choice of \mathcal{C} . This follows since M_2 is independent of M'_2 due to the independent and uniformly distributed key K , and $M_2 \rightarrow M'_2 \rightarrow (M_1, Z^n)$ form a Markov chain. ■

To complete the analysis, we show that $\mathbb{E}[L_1^{(n)}(\mathcal{C})] \rightarrow 0$ as $n \rightarrow \infty$. Recall J is the random variable over $[2^{n(R_C - R')}]$, which shows the index of the picked codeword X^n in the sub-codebook \mathcal{C}_{M_1} . Based on the encoding scheme, J is uniformly distributed and independent of (M_1, \mathcal{C}) . Then it follows

$$\mathbb{E}[L_1^{(n)}(\mathcal{C})]$$

$$\begin{aligned}
&= \frac{1}{n} [I(M_1; Z^n | \mathcal{C})] \\
&= \frac{1}{n} [I(M_1, J; Z^n | \mathcal{C}) - I(J; Z^n | M_1, \mathcal{C})] \\
&\leq \frac{1}{n} [I(X^n; Z^n | \mathcal{C}) - I(J; Z^n | M_1, \mathcal{C})] \tag{46}
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n} [I(X^n; Z^n | \mathcal{C}) - H(J | M_1, \mathcal{C}) + H(J | Z^n, M_1, \mathcal{C})] \\
&= \frac{1}{n} [I(X^n; Z^n | \mathcal{C}) - n(I(X; Z) - \epsilon) + H(J | Z^n, M_1, \mathcal{C})] \tag{47}
\end{aligned}$$

$$\leq \frac{1}{n} [I(X^n; Z^n | \mathcal{C}) - n(I(X; Z) - \epsilon) + n\epsilon_n] \tag{48}$$

$$\leq \frac{1}{n} [I(X^n; Z^n) - nI(X; Z) + n\epsilon + n\epsilon_n] \tag{49}$$

$$= \frac{1}{n} [nI(X; Z) - nI(X; Z) + n\epsilon + n\epsilon_n], \tag{50}$$

which tends to zero as $n \rightarrow \infty$ and $\epsilon \rightarrow 0$. Here

- inequality (46) follows from the fact that $(M_1, J) \rightarrow X^n \rightarrow Z^n$ form a Markov chain,
- equality (47) follows since J is uniformly distributed over $[2^{n(R_c - R')}] = [2^{n(I(X; Z) - \epsilon)}]$ and is independent of (M_1, \mathcal{C}) ,
- inequality follows (48) from Fano's inequality and the fact that $\mathbb{E}[\tilde{P}_e(\mathcal{C})] \rightarrow 0$ as $n \rightarrow \infty$, where

$$\tilde{P}_e(\mathcal{C}_0) = \min_{\tilde{X}^n} \Pr(\tilde{X}^n(Z^n) \neq X^n | \mathcal{C} = \mathcal{C}_0),$$

where the minimization is taken over all functions $\tilde{X}^n : Z^n \rightarrow \mathcal{X}^n$. This can be easily seen if one consider the sub-codebook \mathcal{C}_{M_1} with $2^{n(I(X; Z) - \epsilon)}$ elements as a code for Eve's channel given M_1 is sent, and apply the channel coding theorem [11, Theorem 7.7.1]. Here, we have used the notation ϵ_n to show a sequence such that $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$,

- inequality (49) comes from the fact that $\mathcal{C} \rightarrow X^n \rightarrow Z^n$ form a Markov chain,
- and equality (50) comes from the fact that X_i 's are i.i.d and the channel is memoryless, therefore, $I(X^n; Z^n) = nI(X; Z)$.

To summarize, we showed that $\mathbb{E}[L_1^{(n)}(\mathcal{C})] \rightarrow 0$ as $n \rightarrow \infty$, and hence by Lemma 4, condition (42) holds. Putting (41) and (42) together, we can conclude that there exists at least one codebook which satisfies conditions (9) and (10) simultaneously. This completes the proof.

Remark: This coding scheme can be easily modified to the case in which the feedback channel has a time-invariant rate constraint $\log(|\mathcal{K}_i|) < R_f$ for all i . By using block Markov coding, we can send the key in the L -th block that will be used in the $(L + 1)$ -th block.

V. CONCLUSION

We studied the wiretap channel with a secure rate-limited feedback link and found an upper bound for the secrecy capacity as a function of the feedback rate. The upper bound is achievable in the case of the physically degraded wiretap channel. To achieve the secrecy capacity in this case, Bob ignores the channel output and simply sends back pure randomness, which is used by Alice as a key. To position this

result along Ahlswede and Cai's result [6], suppose that the feedback rate R_f is sufficiently large to send back the entire channel output itself, say, $R_f \geq H(Y)$ (or even $R_f \geq \log |\mathcal{Y}|$). Our result shows that when Bob has an option to choose an arbitrary (stochastic) feedback mapping rather than passively repeating what he has received, the trivial scheme of sending an independently generated secret key is sufficient to achieve the secrecy capacity. In other words, in contrast to the case of [6] where the feedback (output symbols) is only partially useful for a key, the freedom to choose what to send back allows for a full utilization of the feedback data rate R_f . Using the same idea in our scheme it can be shown that for a more capable wiretap channel

$$C_s(R_f) \geq \max_{p(x)} \min \{I(X; Y), R_f + |I(X; Y) - I(X; Z)|^+\},$$

where $|a|^+ = \max\{0, a\}$. In general, this is smaller than our upper bound, and the problem of closing the gap for the wiretap channel models other than (physically) degraded ones remains open for future studies.

REFERENCES

- [1] C. E. Shannon, "Communication theory of secrecy systems," *Bell Syst. Tech. J.*, vol. 28, pp. 656–715, 1949.
- [2] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, pp. 1355–1387, 1975.
- [3] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. IT-24, pp. 339–348, 1978.
- [4] A. El Gamal, "The capacity of a class of broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-25, pp. 166–169, 1979.
- [5] S. K. Leung-Yan-Cheong and M. E. Hellman, "The Gaussian wiretap channel," *IEEE Trans. on Information Theory*, vol. 24, pp. 451–456, 1978.
- [6] R. Ahlswede and N. Cai, "Transmission, identification, and common randomness capacities for wire-tape channels with secure feedback from the decoder," book chapter in *General Theory of Information Transfer and Combinatorics*, LNCS 4123, pp. 258–275, Berlin: Springer-Verlag, 2006.
- [7] H. Yamamoto, "Rate-distortion theory for the Shannon cipher system," *IEEE Trans. Inf. Theory*, vol. IT-43, pp. 827–835, 1997.
- [8] N. Merhav, "Shannon's secrecy system with informed receivers and its application to systematic coding for wiretapped channel," in *Proc. Internat. Symp. Inf. Theory*, Nice, France, 2007.
- [9] L. Lai, H. El Gamal, and V. Poor, "The wiretap channel with feedback: encryption over the channel," *IEEE Trans. Inf. Theory*, vol. 54, pp. 5059–5067, 2008.
- [10] E. Tekin and A. Yener, "The general Gaussian multiple access channel and two-way wire-tap channels: Achievable rates and cooperative jamming," submitted to *IEEE Trans. Inf. Theory*, Feb 2007.
- [11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.