Wiretap Channel with Rate-limited Feedback

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

***Abstract*—This paper studies the problem of secure communication over a degraded wiretap channel $p(y,z|x) = p(y|x)p(z|y)$ with 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 wiretapper respectively. The secrecy capacity is characterized as**

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

A capacity-achieving coding scheme is presented, in which the receiver securely feeds back fresh randomness with rate R_f, *independent* of the received channel output. The transmitter then uses the shared randomness as a secret key on top of Wyner's coding scheme for wiretap channel without feedback. Hence, when the receiver has a means of interacting with the transmitter, he should allocate all resources to convey a new key rather than sending back the channel output. For the converse, a recursive argument is used to obtain the single-letter characterization.

I. Introduction

Shannon [1] studied a secrecy system consisting of a legitimate transmitter (Alice), a legitimate receiver (Bob), and an eavesdropper (Eve), in which Alice wishes to transmit a message W to Bob completely secret from Eve. He showed that if Eve has access to what Bob receives, a secret key K whose entropy satisfies $H(K) \geq H(W)$ has to be shared between Alice and Bob. Later, 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 $p(z|y)$; see Figure 1.

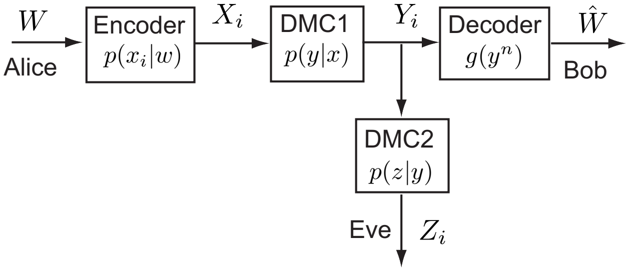

Fig. 1. Degraded wiretap channel.

He showed that Alice can exploit the better quality of her channel to Bob and transmit information securely at a positive rate

$$C_s = \max_{p(x)}[I(X;Y) - I(X;Z)] = \max_{p(x)} I(X;Y|Z), \qquad (1)$$

even without any secret key. Later, this exciting result was extended by Csiszár and Körner [3] to general broadcast channels with confidential messages. In particular, they showed that the secrecy capacity C_s, the supremum of all achievable rates of secure communication, is strictly positive unless the channel from Alice to Eve is *less noisy* [4] than the channel from Alice to Bob.

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 Wyner's wiretap channel model by introducing a secure feedback link of rate R_f from Bob to Alice as depicted in Figure 2. There are several concrete scenarios where this model is applicable. For instance, consider the communication between a satellite (Alice) and a base station (Bob) on the ground. The satellite broadcasts information to the ground, so any (unintended) station can wiretap it. The base station can beamform some data back to the satellite securely, which can be used to enhance the secret data rate sent from the satellite to the base station.

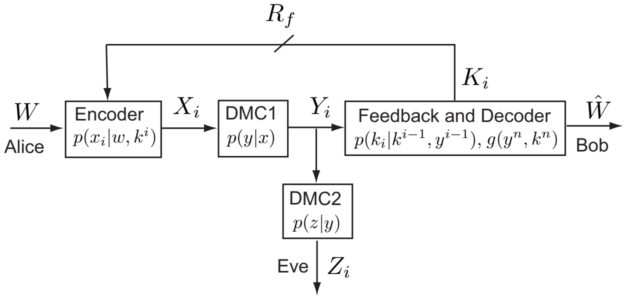

Fig. 2. Degraded wiretap channel with secure rate-limited feedback.

The main purpose of this paper is to characterize the secrecy capacity $C_s(R_f)$ as a function of the secure feedback rate R_f. We show that the secrecy capacity is given by

$$C_s(R_f) = \max_{p(x)} \min[I(X;Y),\ I(X;Y|Z) + R_f]. \qquad (2)$$

Interestingly, we show that to achieve the secrecy capacity Bob can simply ignore what he receives and sends "fresh" randomness. This fresh randomness plays the role of a secret

978-1-4244-2571-6/08/$25.00 ©2008 IEEE

key, which bridges Shannon's original result and Wyner's wiretap model. We modified Wyner's original scheme to allow the use of a shared key (sent from Bob to Alice via rate-limited feedback). To be fair, this modification has been already proposed by Yamamoto [6] and Merhav [7], who characterized the secrecy capacity of wiretap channels with shared key (already given prior to the communication), where additional effects of having distortion or side information are also considered.

Proving the optimality of (2) is a little more involved. Due to the dependencies introduced by the feedback, we use a *recursive* argument to obtain the desired single-letter characterization. Exploiting the recursive structure to find the single-letter characterization could be a powerful tool for similar converse proofs.

In a closely related work, Ahlswede and Cai [5] studied degraded wiretap channels with secure *output* feedback, in which channel output symbols received by Bob are secretly fed back to Alice. They showed that the secrecy capacity is given by

$$C_{sf} = \max_{p(x)} \min\{I(X;Y),\ I(X;Y|Z) + H(Y|X,Z)\} \quad (3)$$

which results in capacity larger than the non-feedback case (1). At a first glance, it seems contradictory that the optimal receiver should ignore the channel outputs completely when feedback is rate-limited (our paper) while the unlimited output feedback (Ahlswede–Cai) boosts the secrecy capacity as in (3). However, looking closer at the coding scheme by Ahlswede and Cai, one realizes that their scheme essentially extracts the channel randomness hidden in those outputs and uses that 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, thorough studies have been conducted on characterizing the secrecy capacity of various two-way communication systems. Lai, El Gamal, and Poor [8] 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, Alice can send messages securely at the capacity of her channel to Bob. Tekin and Yener [9] presented an achievable region for two-way Gaussian wiretap channels. In their model, Eve also receives the sum of the signals from both transmitters, and a Gaussian noise. They showed that due to the multiple access nature of Eve's channel, both transmitters can help to hide the other user's message while maintaining the communication rate. In both studies, the additive nature of Eve's channel gives the opportunity for jamming, in addition to possible backward information transfer. In other words, the feedback provides rather "physical" advantages for Alice–Bob interactions. In comparison, our model decouples the forward and backward communication channels, eliminating the possible use of jamming, and focuses on rather "inherent" advantages of Alice–Bob interactions. 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 converse and the achievability scheme in Section 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 $W \in [2^{nR}] := \{1, 2, \ldots, 2^{nR}\}$ over a degraded broadcast channel $p(y, z|x) = p(y|x)p(z|y)$, 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 wiretapper Eve. Alice and Bob wish to communicate the message W reliably over their channel $p(y|x)$ while keeping it secret from Eve. To enhance the secrecy of their communication, Bob can communicate back signals $K_i \in \mathcal{K}_i$, $i = 1, 2, \ldots, n$, over a feedback link of rate R_f secret from Eve. The feedback signal K_i at time i can depend causally on previous channel outputs $Y^{i-1} := (Y_1, \ldots, Y_{i-1})$ and previous feedback signals K^{i-1}. We assume that the channel alphabets $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$, and feedback alphabets $\mathcal{K}_1, ..., \mathcal{K}_n$ are finite, and Eve has complete knowledge about them as well as the feedback-encoding scheme used by Alice and Bob. The broadcast channel $p(y, z|x)$ is memoryless and degraded, i.e.,

$$\begin{aligned} p(y_i, z_i|x^i, y^{i-1}, z^{i-1}) &= p(y_i, z_i|x_i) \\ &= p(y_i|x_i)p(z_i|y_i) \end{aligned}$$

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

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

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

$$\frac{1}{n}\sum_{i=1}^{n} \log(|\mathcal{K}_i|) \le R_f, \quad (4)$$

2) stochastic encoding maps consisting of conditional probability distributions $f_i(x_i|w, k^i)$, $i = 1, 2, \ldots, n$, defined for each $x_i \in \mathcal{X}$, $k^i \in \mathcal{K}^i := \mathcal{K}_1 \times \cdots \times \mathcal{K}_i$, and $w \in [2^{nR}]$ such that for each i, w, k^i, $\sum_{x_i} f_i(x_i|w, k^i) = 1$ (in other words, $f_i(x_i|w, k^i)$ denotes the probability that at time i the message w and the previously received feedback signals k^i are mapped to the channel input x_i),
3) stochastic feedback maps consisting of conditional probability distributions $\psi_i(k_i|y^{i-1}, k^{i-1})$ (by convention, $K_1 \sim \psi_1(k_1)$ independent of W), and
4) a decoding map $g : \mathcal{Y}^n \times \mathcal{K}^n \to [2^{nR}]$ resulting in the decoded message

$$\hat{W} = g(Y^n, K^n). \quad (5)$$

We assume throughout that the message W is a random variable uniformly distributed over $[2^{nR}]$. The probability of error is defined as $P_e^{(n)} = \Pr\{\hat{W} \neq W\}$.

Definition 2.1: We define the equivocation $\Delta^{(n)}$ as

$$\Delta^{(n)} = \frac{H(W|Z^n)}{H(W)} = \frac{H(W|Z^n)}{nR}.$$

Definition 2.2: We call the pair (R, d) achievable if there exists a sequence of $(2^{nR}, 2^{nR_f}, n)$ codes such that

$$P_e^{(n)} \to 0, \tag{6}$$

$$\Delta^{(n)} \to d, \tag{7}$$

as $n \to \infty$. We call rate R *secret* if the pair $(R, 1)$ is achievable.

Definition 2.3: We define the rate–equivocation region $\mathcal{R} = \mathcal{R}(R_f)$ as the closure of all achievable pair (R, d).

Definition 2.4: We define the secrecy capacity $C_s(R_f)$ at feedback rate R_f as

$$C_s(R_f) = \sup_{(R,1)\in\mathcal{R}(R_f)} R.$$

We are ready to state our main result.

Theorem 2.1: The secrecy capacity of a degraded wiretap channel with rate-limited feedback R_f is

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

Roughly speaking, until saturated by the mutual information $I(X;Y)$, adding one bit secure feedback can increase the forward secure rate by one bit. Note that this does not necessarily mean $C_s(R_f) = \min\{C_s(0) + R_f,\ C_m\}$, where $C_m = \max_{p(x)} I(X;Y)$ is the capacity of the main channel from Alice to Bob; it is easy to find examples, in which $C_s(R_f) < C_s(0) + R_f \le C_m$.

We prove the converse in the next section. A capacity-achieving coding scheme is presented in Section IV.

III. Proof of the converse

In this section we show if the pair $(R, 1)$ is achievable, R must satisfy the following constraint:

$$R \le \max_{p(x)} \min[I(X;Y),\ I(X;Y|Z) + R_f]. \tag{8}$$

To show (8) we prove the following two upper bounds:

$$R \le \frac{1}{n}\sum_{i=1}^{n} I(X_i;Y_i). \tag{9}$$

$$R \le R_f + \frac{1}{n}\sum_{i=1}^{n} I(X_i;Y_i|Z_i), \tag{10}$$

Then we can use the usual time sharing random variable and concavity of mutual information in $p(x)$ to obtain (8).

First, (9) follows easily from Fano's inequality. For details, please refer to the the standard converse proof of the channel coding theorem [10, Theorem 7.7.1].

We now prove (10) using Fano's inequality, secrecy constraint (7), and rate-limit constraint (4). A recursive argument (Lemma 3.1) is used to obtain the single-letter characterization.

By Fano's inequality we have

$$H(W|\hat{W}) \le 1 + P_e^{(n)} nR \triangleq n\epsilon_n,$$

where $\epsilon_n \to 0$ as $n \to \infty$ by the assumption that $P_e^{(n)} \to 0$. From (5) and data processing inequality we have

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

By definition of $\Delta^{(n)}$, we have

$$\begin{aligned} I(W;Z^n) &= H(W) - H(W)\Delta^{(n)} \\ &= nR(1-\Delta^{(n)}) = n\gamma_n, \end{aligned} \tag{11}$$

where $\gamma_n \to 0$ as $n \to \infty$ by the assumption that $\Delta^{(n)} \to 1$. It then follows that

$$\begin{aligned} nR &= H(W) \\ &= H(W|Z^n) + I(W;Z^n) \\ &= H(W|Z^n) + n\gamma_n && (12) \\ &= I(W;Y^n,K^n|Z^n) + H(W|Y^n,Z^n,K^n) + n\gamma_n \\ &\le I(W;Y^n,K^n|Z^n) + n\epsilon_n + n\gamma_n && (13) \\ &= I(W;K^n|Z^n) + I(W;Y^n|K^n,Z^n) + n\delta_n && (14) \\ &\le H(K^n|Z^n) + I(W;Y^n|K^n,Z^n) + n\delta_n, && (15) \end{aligned}$$

where (12) follows from (11), (13) follows from Fano's inequality, (14) follows by defining $\delta_n = \epsilon_n + \gamma_n$, and (15) follows from the fact that entropy is positive.

The following lemma is crucial to single-letterize (15).

Lemma 3.1: For each $j = 1, 2, \dots, n$, we have

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

Proof: We have the following chain of inequalities:

$$\begin{aligned} &H(K^j|Z^j) + I(W;Y^j|K^j,Z^j) \\ &= H(K^j|Z^j) + I(W;Y^{j-1}|K^j,Z^j) \\ &\quad + I(W;Y_j|Y^{j-1},K^j,Z^j) \\ &\le H(K^j|Z^j) + I(W;Y^{j-1}|K^j,Z^j) \\ &\quad + I(W,Y^{j-1},K^j,Z^{j-1},X_j;Y_j|Z_j) && (16) \\ &= H(K^j|Z^j) + I(W;Y^{j-1}|K^j,Z^j) + I(X_j;Y_j|Z_j) && (17) \\ &\le H(K^j|Z^j) + I(W,Z_j;Y^{j-1}|K^j,Z^{j-1}) + I(X_j;Y_j|Z_j) && (18) \\ &= H(K^j|Z^j) + I(W;Y^{j-1}|K^j,Z^{j-1}) + I(X_j;Y_j|Z_j) && (19) \\ &= H(K^j|Z^j) + I(W,K_j;Y^{j-1}|K^{j-1},Z^{j-1}) \\ &\quad - I(K_j;Y^{j-1}|K^{j-1},Z^{j-1}) + I(X_j;Y_j|Z_j) \\ &= H(K^j|Z^j) + I(W;Y^{j-1}|K^{j-1},Z^{j-1}) + I(X_j;Y_j|Z_j) \\ &\quad + I(K_j;Y^{j-1}|W,K^{j-1},Z^{j-1}) \\ &\quad - I(K_j;Y^{j-1}|K^{j-1},Z^{j-1}) \\ &= H(K^{j-1}|Z^j) + I(W;Y^{j-1}|K^{j-1},Z^{j-1}) + I(X_j;Y_j|Z_j) \end{aligned}$$

$$\begin{aligned}
&\quad + I(K_j; Y^{j-1}|W, K^{j-1}, Z^{j-1}) + H(K_j|K^{j-1}, Z^j) \\
&\quad + H(K_j|Y^{j-1}, K^{j-1}, Z^{j-1}) - H(K_j|K^{j-1}, Z^{j-1})
\end{aligned}$$

$$\begin{aligned}
&\leq H(K^{j-1}|Z^j) + I(W; Y^{j-1}|K^{j-1}, Z^{j-1}) + I(X_j; Y_j|Z_j) \\
&\quad + I(K_j; Y^{j-1}|W, K^{j-1}, Z^{j-1}) \\
&\quad + H(K_j|Y^{j-1}, K^{j-1}, Z^{j-1}) \qquad (20) \\
&\leq H(K^{j-1}|Z^j) + I(W; Y^{j-1}|K^{j-1}, Z^{j-1}) + I(X_j; Y_j|Z_j) \\
&\quad + H(K_j|W, K^{j-1}, Z^{j-1}) + H(K_j|Y^{j-1}, K^{j-1}, Z^{j-1}) \\
&\quad - H(K_j|Y^{j-1}, W, K^{j-1}, Z^{j-1}) \\
&= H(K^{j-1}|Z^j) + I(W; Y^{j-1}|K^{j-1}, Z^{j-1}) + I(X_j; Y_j|Z_j) \\
&\quad + H(K_j|W, K^{j-1}, Z^{j-1}) \qquad (21) \\
&\leq H(K^{j-1}|Z^{j-1}) + I(W; Y^{j-1}|K^{j-1}, Z^{j-1}) \\
&\quad + H(K_j|W, K^{j-1}, Z^{j-1}) + I(X_j; Y_j|Z_j), \qquad (22)
\end{aligned}$$

where

- (16) and (18) follow from the fact that the mutual information is positive,
- (17) holds because the channel is memoryless and therefore $Y_j \to (X_j, Z_j) \to (W, Y^{j-1}, K^j, Z^{j-1})$ form a Markov chain,
- (19) holds because $Z_j \to (W, K^j, Z^{j-1}) \to Y^{j-1}$ form a Markov chain,
- (20) and (22) follow from the fact that conditioning reduces entropy, and
- (21) holds because of the following Markov chain $(W, Z^{j-1}) \to (Y^{j-1}, K^{j-1}) \to K_j$. ■

The following corollary follows immediately from Lemma 3.1 by induction.

Corollary 3.2: We have

$$\begin{aligned}
&H(K^n|Z^n) + I(W; Y^n|K^n, Z^n) \\
&\quad \leq \sum_{i=1}^{n} I(X_i; Y_i|Z_i) + H(K_i|W, K^{i-1}, Z^{i-1}).
\end{aligned}$$

Applying Corollary 3.2 to (15) we obtain

$$\begin{aligned}
R &\leq \frac{1}{n}\sum_{i=1}^{n} I(X_i; Y_i|Z_i) + H(K_i|W, K^{i-1}, Z^{i-1}) + \delta_n \\
&\leq \frac{1}{n}\sum_{i=1}^{n} I(X_i; Y_i|Z_i) + \frac{1}{n}\sum_{i=1}^{n} H(K_i) + \delta_n \\
&\leq \frac{1}{n}\sum_{i=1}^{n} I(X_i; Y_i|Z_i) + R_f + \delta_n, \qquad (23)
\end{aligned}$$

where inequality (23) follows from the rate-limit constraint (4). Finally, letting $n \to \infty$ and $\delta_n \to 0$, we get (10).

To complete the proof, let Q be a time-sharing random variable distributed uniformly over $\{1, 2, ..., n\}$ and independent of X^n, Y^n, Z^n. Then (10) can be written as

$$\begin{aligned}
R &\leq R_f + \frac{1}{n}\sum_{i=1}^{n} I(X_i; Y_i|Z_i) \\
&= R_f + \frac{1}{n}\sum_{i=1}^{n} I(X_i; Y_i|Z_i, Q = i) \\
&= R_f + I(X_Q; Y_Q|Z_Q, Q) \\
&= R_f + I(X; Y|Z, Q), \qquad (24)
\end{aligned}$$

where $X \triangleq X_Q$, $Y \triangleq Y_Q$, $Z \triangleq Z_Q$. Similarly, (9) can be written as

$$R \leq I(X; Y|Q). \qquad (25)$$

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. Thus $Q \to X \to Y \to Z$ form a Markov chain so that we have from (24)

$$\begin{aligned}
R &\leq R_f + I(X; Y|Z, Q) \\
&\leq R_f + I(X, Q; Y|Z) \\
&= R_f + I(X; Y|Z), \qquad (26)
\end{aligned}$$

and similarly from (25)

$$R \leq I(X; Y|Q) \leq I(X, Q; Y) = I(X; Y). \qquad (27)$$

Combining (26) and (27) 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 converse.

IV. Capacity-achieving coding scheme

In this section we present a coding scheme that acheives

$$R = \max_{p(x)} \min\ [I(X; Y),\ I(X; Y|Z) + R_f]. \qquad (28)$$

Consider a wiretap channel where a secret key of rate R_k is shared between Alice and Bob. The model is shown in Figure 3. Let $\mathcal{R}'(R_k)$ be the equivocation–rate region of this model. If we set $R_k = R_f$ then we can easily see $\mathcal{R}'(R_k) \subseteq \mathcal{R}(R_f)$, where $\mathcal{R}(R_f)$ is the equivocation–rate region with secure feedback rate-limited by R_f (recall Figure 2). This is true because Bob can use the feedback link to send back a secret key of rate R_f (i.e., send K_1 with $|\mathcal{K}_1| = 2^{nR_f}$). Hence, we can use the coding scheme for the wiretap channel with shared key to achieve the secrecy rate (28).

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

Details of the coding scheme are as follows.

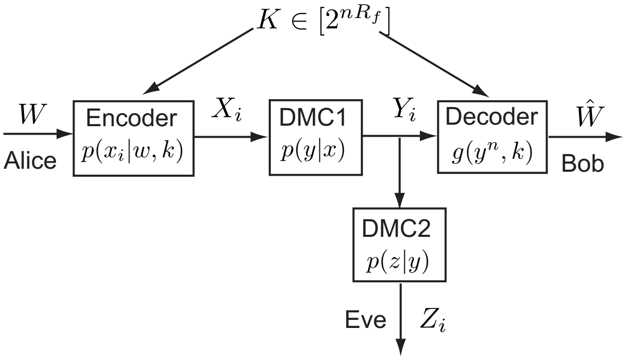

Fig. 3. Wiretap channel with shared key.

Fix any $p(x)$ and let $R'_f \le R_f$ be such that

$$R = \min[I(X;Y),\ I(X;Y|Z) + R_f] = I(X;Y|Z) + R'_f. \tag{29}$$

Let $R' = R - R'_f = I(X;Y|Z)$, and $W = \{W_1, W_2\}$, where W_1 and W_2 are independent random variables uniformly distributed over $\mathcal{W}_1 = [2^{nR'}]$ and $\mathcal{W}_2 = [2^{nR'_f}]$, respectively. The message W_1 will be transmitted securely using Wyner's original coding scheme while the secrecy of W_2 will be achieved by using the key of rate R'_f.

Codebook generation. Generate a codebook consisting of $2^{n(I(X;Y)-\epsilon)}$ i.i.d. codewords X^n according to probability distribution $\prod_i^n p(x_i)$. Divide the codebook into $2^{nR'}$ disjoint *sub-codebooks*, each of which has $2^{n(I(X;Y)-R'-\epsilon)}$ codewords. Label the sub-codebooks $C_1, ..., C_{2^{nR'}}$. Now, divide each sub-codebook C_i into $T = 2^{n(I(X;Y)-R'-R'_f-\epsilon)}$ *sections* $C_{i1}, ..., C_{iT}$, each of which has $2^{nR'_f}$ codewords. Enumerate the codewords in each section from 1 to $2^{nR'_f}$.

Feedback. Let K be uniform over $\mathcal{W}_2 = [2^{nR'_f}]$. Bob at time 1, sends $K_1 = K$.

Encoding. We use K as a key shared between Alice and Bob. Generate a new variable $\tilde{W}_2 = W_2 \oplus K \in [2^{nR'_f}]$, where $\oplus$ is the modulo addition over the set $\mathcal{W}_2 = [2^{nR'_f}]$. Note that K and W_2 are uniformly distributed and independent, so $\tilde{W}_2$ is uniformly distributed and independent of both K and W_2.

We pick $X^n(W_1, W_2)$ as follows. According to W_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 $\tilde{W}_2$ we pick the corresponding codeword among the $2^{nR'_f}$ codewords in that section.

Decoding. Upon receiving Y^n decode $\hat{X}^n = D(Y^n)$. Decode W_1 to $\hat{W}_1$, the index of the sub-codebook $\hat{X}^n$ belongs to. Decode $\tilde{W}_2$ to $\hat{\tilde{W}}_2$, the index of $\hat{X}$ in the section it belongs to, and then take $\hat{W}_2 = \hat{\tilde{W}}_2 \ominus K$, where $\ominus$ is the modulo subtraction over $\mathcal{W}_2$.

Analysis of error probability and secrecy analysis. Since there are $2^{n(I(X;Y)-\epsilon)}$ codewords, it is easy to check that there exists a codebook with vanishing decoding error probability [10, Theorem 7.7.1]. Also note that the probability distribution on the sub-codebooks is uniform, because the subsections are chosen uniformly and the codeword in each subsection is picked according to the key which is also a uniform random variable. This is the main point in Wyner's analysis in [2]. Therefore, the secrecy analysis can be done in parallel to Wyner's original analysis. Details are omitted.

V. Conclusion

We studied the wiretap channel with secure rate-limited feedback link and characterized the secrecy capacity as a function of the feedback rate. To achieve the secrecy capacity, 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 [5], suppose that the feedback rate R_f is sufficiently large to send back the entire channel output itself, say, $R_f \ge H(Y)$ (or even $R_f \ge \log|\mathcal{Y}|$). Our result proves 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 [5] 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.

Ultimately, we are interested in characterizing the secrecy capacity region when information flows in both directions over the two-way communication channels. To get some insight, we considered a special case of the secrecy capacity from Alice to Bob when the backward channel is secure, rate-limited, and independent of the forward channel. Future studies will focus on characterizing the secrecy capacity region for different scenarios of two-way communication.

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] J. Körner and K. Marton, "Comparison of two noisy channels," in *Topics in Information Theory*, Keszthely (Hungry) 1975, Colloquia Math. Soc. Janos Bolyai, pp. 411–423, Amsterdam: North-Holland, 1977,

[5] 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.

[6] H. Yamamoto, "Rate-distortion theory for the Shannon cipher system" *IEEE Trans. Inf. Theory*, vol. IT-43, pp. 827–835, 1997.

[7] 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.

[8] L. Lai, H. El Gamal, and V. Poor, "The wiretap channel with feedback: encryption over the channel," submitted to *IEEE Trans. Inf. Theory*, Apr. 2007.

[9] 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.

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