entropy

MDPI

Article

How Can We Fully Use Noiseless Feedback to Enhance the Security of the Broadcast Channel with Confidential Messages

Xin Li [1,2,†,‡], Bin Dai [1,2,*,†,‡] and Zheng Ma [1,†,‡]

1 School of Information Science and Technology, Southwest JiaoTong University, Chengdu 611756, China; 15884501083@163.com (X.L.); zma@home.swjtu.edu.cn (Z.M.)
2 The State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China
* Correspondence: daibin@home.swjtu.edu.cn; Tel.: +86-135-480-53724
† Current address: School of Information Science and Technology, Southwest JiaoTong University, Chengdu 611756, China
‡ These authors contributed equally to this work.

Received: 1 August 2017; Accepted: 2 October 2017; Published: 6 October 2017

Abstract: The model for a broadcast channel with confidential messages (BC-CM) plays an important role in the physical layer security of modern communication systems. In recent years, it has been shown that a noiseless feedback channel from the legitimate receiver to the transmitter increases the secrecy capacity region of the BC-CM. However, at present, the feedback coding scheme for the BC-CM only focuses on producing secret keys via noiseless feedback, and other usages of the feedback need to be further explored. In this paper, we propose a new feedback coding scheme for the BC-CM. The noiseless feedback in this new scheme is not only used to produce secret keys for the legitimate receiver and the transmitter but is also used to generate update information that allows both receivers (the legitimate receiver and the wiretapper) to improve their channel outputs. From a binary example, we show that this full utilization of noiseless feedback helps to increase the secrecy level of the previous feedback scheme for the BC-CM.

Keywords: Broadcast channel with confidential messages; noiseless feedback; secrecy capacity region; source coding with side information

1. Introduction

Wyner, in his outstanding paper on the degraded wiretap channel [1], first studied secure transmission over a physically degraded broadcast channel in the presence of an additional wiretapper. Wyner showed that the secrecy capacity (the maximum transmission rate with perfect secrecy constraint) of the degraded wiretap channel model was given by

$$C_s^d = \max_{P(x)} (I(X;Y) - I(X;Z)), \tag{1}$$

where X, Y and Z are the channel input, channel output for the legitimate receiver and channel output for the wiretapper, respectively, and they satisfy the Markov chain $X \to Y \to Z$. Note here that the secrecy capacity defined in (1) can be viewed as the difference between the main channel capacity $I(X;Y)$ (the channel for the transmitter and the legitimate receiver) and the wiretap channel capacity $I(X;Z)$ (the channel for the transmitter and the wiretapper). Later, Csiszár and Körner [2] extended Wyner's work [1] to a more general case: the broadcast channel with confidential messages (BC-CM), where common and confidential messages were transmitted through a discrete memoryless general broadcast channel (without the degradedness assumption $X \to Y \to Z$), and the common message

was intended to be decoded by both the legitimate receiver and the wiretapper, while the confidential message was only allowed to be decoded by the legitimate receiver. The secrecy capacity region (the capacity region with the perfect secrecy constraint) of this generalized model is determined in [2], and it is given by

$$\begin{aligned} \mathcal{C}_s = \{(R_0, R_1) : 0 \leq R_0 \leq \min\{I(U;Y), I(U;Z)\} \\ 0 \leq R_1 \leq I(V;Y|U) - I(V;Z|U)\}, \end{aligned} \tag{2}$$

where U and V respectively represent the common message and the confidential message, and R_0 and R_1 are the transmission rates of the common message and the confidential message, respectively. Here note that from (2), it is not difficult to show that the secrecy capacity C_s (the maximum transmission rate of the confidential message with the perfect secrecy constraint) of the BC-CM is given by

$$C_s = \max_{P(v,x)} [I(V;Y) - I(V;Z)]^+, \tag{3}$$

where the function $[x]^+ = x$ if $x \geq 0$, else $[x]^+ = 0$, and C_s is also called the secrecy capacity of the general wiretap channel. The work of [1] and [2] lays the foundation of the physical layer security in modern communication systems.

Recently, Ahlswede and Cai [3] found that if the legitimate receiver sent his own channel output Y back to the transmitter through a noiseless feedback channel, the secrecy capacity region $\mathcal{C}_s$ of the BC-CM could be expanded to an achievable secrecy rate region

$$\begin{aligned} \mathcal{C}_s^{f-cai} = \{(R_0, R_1) : 0 \leq R_0 \leq \min\{I(U;Y), I(U;Z)\} \\ 0 \leq R_1 \leq \min\{[I(V;Y|U) - I(V;Z|U)]^+ + H(Y|Z,U,V), I(V;Y|U)\}\}, \end{aligned} \tag{4}$$

where the auxiliary random variables U and V are defined similarly as those in (2). The coding scheme of the region $\mathcal{C}_s^{f-cai}$ combines Csiszár and Körner's coding scheme for the BC-CM [2] with the idea of using a secret key to encrypt the transmitted message, where the secret key is generated from the noiseless feedback. Note here that the region $\mathcal{C}_s^{f-cai}$ is an inner bound on the secrecy capacity $\mathcal{C}_s^f$ of the BC-CM with noiseless feedback, and to the best of the authors' knowledge, C_s^f remains unknown. Similar to the work of [2], using (4), Ahlswede and Cai also provided an achievable secrecy rate R_s^{f-cai} (lower bound on the secrecy capacity) of the general wiretap channel with noiseless feedback, and it is given by

$$R_s^{f-cai} = \max_{P(v,x)} \min\{[I(V;Y) - I(V;Z)]^+ + H(Y|V,Z), I(V;Y)\}, \tag{5}$$

where V is defined in the same way as in (2). In [3], Ahlswede and Cai further pointed out that for the degraded wiretap channel with noiseless feedback (the Markov chain $X \to Y \to Z$ holds), the secrecy capacity C_s^{df} was given by

$$C_s^{df} = \max_{P(x)} \min\{I(X;Y) - I(X;Z) + H(Y|X,Z), I(X;Y)\}. \tag{6}$$

Here, note that the secrecy capacities in (5) and (6) can be viewed as a combination of two parts: the first part is the difference between the main channel capacity ($I(V;Y)$ or $I(X;Y)$) and the wiretap channel capacity ($I(V;Z)$ or $I(X;Z)$), and the second part is the rate $H(Y|V,Z)$ ($H(Y|X,Z)$) of a secret key generated by the noiseless feedback and shared between the legitimate receiver and the transmitter. Comparing (6) with (1) and (5) with (3), it is easy to see that by using the noiseless feedback to generate a secret key encrypting the transmitted message, the secrecy capacity of the wiretap channel can

be enhanced. Besides the work of [3], other related works on the BC-CM or wiretap channel in the presence of noiseless feedback are in [4–7].

In this paper, we re-visit the BC-CM with noiseless feedback investigated by Ahlswede and Cai [3] (see Figure 1), and we propose a new achievable secrecy rate region for this feedback model. The coding scheme for this achievable region combines the previous Ahlswede and Cai's scheme [3] with the Wyner-Ziv scheme for lossy source coding with side information [8], i.e., compared with Ahlswede and Cai's scheme, in our new scheme, the noiseless feedback is not only used to produce the secret key but also used to generate an update information that allows the legitimate receiver to improve his channel output. From a binary example, we show that this full utilization of noiseless feedback helps to obtain a larger achievable secrecy rate of the confidential message.

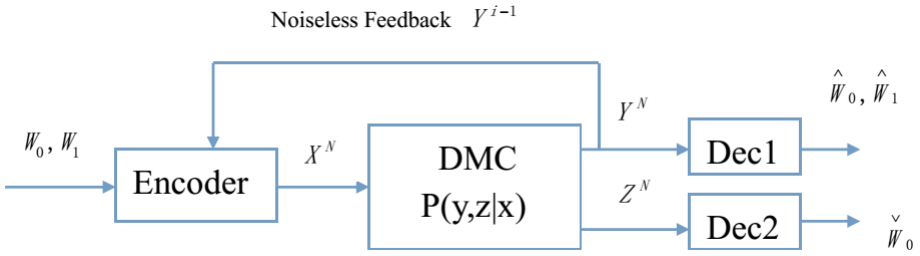

Figure 1. Broadcast channel with confidential messages and noiseless feedback.

Now the remainder of this paper is organized as follows. Section 2 is about the problem formulation and the main result of this paper. A binary example is provided in Section 3. Final conclusions are presented in Section 4.

2. Problem Formulation and New Result

Notations: In this paper, random variables are written in upper case letters (e.g. V), real values are written in lower case letters (e.g. v), and members of the alphabet are written in calligraphic letters (e.g. $\mathcal{V}$). Random vectors and their values are written in a similar way. The probability $Pr\{V = v\}$ is shortened to $P(v)$. In addition, for the remainder of this paper, the base of the logarithm is 2.

Model description: Suppose that the common message W_0 is chosen to be transmitted, and it is uniformly distributed over its alphabet $\mathcal{W}_0 = \{1, 2, ..., M_0\}$. Analogously, the confidential message W_1 is chosen to be transmitted, and it is uniformly distributed over its alphabet $\mathcal{W}_1 = \{1, 2, ..., M_1\}$. The channel is discrete and memoryless with input X^N, outputs Y^N, Z^N, and has transition probability $P(y, z|x)$. At time i ($1 \leq i \leq N$), the legitimate receiver receives the channel output Y_i, and he sends the previous channel outputs Y_1,...,Y_{i-1} back to the transmitter via a noiseless feedback channel. Hence at time i, the channel encoder f_i is denoted by

$$X_i = \begin{cases} f_i(W_0, W_1), & i = 1 \\ f_i(W_0, W_1, Y^{i-1}), & 2 \leq i \leq N. \end{cases} \tag{1}$$

Here we should note that f_i does not need to be deterministic and stochastic encoding is also allowed. For the legitimate receiver, after receiving Y^N, he uses a decoding mapping $\psi_1 : \mathcal{Y}^N \to \mathcal{W}_0 \times \mathcal{W}_1$, to obtain $\hat{W}_0$ and $\hat{W}_1$, which are estimations of the transmitted messages W_0 and W_1, respectively. The legitimate receiver's decoding error probability P_{e1} is defined by

$$P_{e1} = \frac{1}{M_0 M_1} \sum_{i=1}^{M_0} \sum_{j=1}^{M_1} Pr\{\psi_1(y^N) \neq (i, j) | (i, j) \text{ sent}\}. \tag{2}$$

For the wiretapper, after receiving Z^N, he uses a decoding mapping $\psi_2 : \mathcal{Z}^N \to \mathcal{W}_0$, to obtain $\check{W}_0$, which is an estimation of the transmitted message W_0. Moreover, the wiretapper also tries to decode

the transmitted message W_1 via his own channel output Z^N, and his equivocation (uncertainty) about W_1 is denoted by

$$\Delta = \frac{1}{N} H(W_1|Z^N). \tag{3}$$

The wiretapper's decoding error probability P_{e2} is defined by

$$P_{e2} = \frac{1}{M_0} \sum_{i=1}^{M_0} Pr\{\psi_2(z^N) \neq i | i \text{ sent}\}. \tag{4}$$

Finally, using similar criteria in [1] and [2], if for any small positive number ϵ, there exists an encoding-decoding scheme with parameters M_0, M_1, N, P_{e1} and P_{e2} such that

$$\frac{\log M_0}{N} \geq R_0 - \epsilon, \ \frac{\log M_1}{N} \geq R_1 - \epsilon, \ \Delta \geq R_1 - \epsilon, \ P_{e1} \leq \epsilon, \ P_{e2} \leq \epsilon, \tag{5}$$

we say that the rate pair (R_0, R_1) is achievable with perfect secrecy. The secrecy capacity region C_s^f is composed of all achievable secrecy rate pairs satisfying (5), and the following Theorem 1 provides an inner bound on C_s^f.

Theorem 1. *The secrecy capacity region C_s^f of the discrete memoryless BC-CM with noiseless feedback satisfies*

$$C_s^f \supseteq C_s^{f-new}, \tag{6}$$

where

$$\begin{aligned} C_s^{f-new} = \{(R_0, R_1) : 0 \leq R_0 \leq \min\{I(U;Y,V_1), I(U;Z,V_2)\} - I(U,V,Y;V_0,V_2|Z) \\ 0 \leq R_1 \leq \min\{[I(V;Y,V_1|U) - I(V;Z,V_2|U)]^+ + H(Y|Z,U,V,V_2), I(V;Y,V_1|U)\} \\ 0 \leq R_0 + R_1 \leq \min\{I(U;Y,V_1), I(U;Z,V_2)\} + I(V;Y,V_1|U) - I(V_1;U,V,Y|V_0,Y) \\ -I(V_2;U,V,Y|V_0,Z) - \max\{I(V_0;U,V,Y|Y), I(V_0;U,V,Y|Z)\}\}, \end{aligned}$$

the joint probability mass function $P(v_0, v_1, v_2, u, v, x, y, z)$ is denoted by

$$P(v_0, v_1, v_2, u, v, x, y, z) = P(v_0, v_1, v_2|u, v, y) P(y, z|x) P(x|u, v) P(v|u) P(u), \tag{7}$$

and the auxiliary random variables V_0, V_1, V_2, V, U take values in finite alphabets.

Proof. The coding scheme for the inner bound C_s^{f-new} combines the previous Ahlswede and Cai's scheme of the model of Figure 1 with a "generalized" Wyner-Ziv scheme for lossy source coding with side information [8], and the details of the proof of Theorem 1 are in Appendix A. □

Remark 1. *There are some notes on Theorem 1; see the following.*

- *Comparing our new inner bound C_s^{f-new} with the previous Ahlswede and Cai's inner bound C_s^{f-cai}, in general, we do not know which one is larger. In the next section, we consider a binary case of the BC-CM with noiseless feedback, and compute these inner bounds for this binary case. From this binary example, we show that the maximum achievable R_1 (the transmission rate of the confidential message with perfect secrecy constraint) in C_s^{f-new} is larger than that in C_s^{f-cai}, however, the enhancement of R_1 is at the cost of reducing the transmission rate of the common message R_0.*
- *Note here that in C_s^{f-new}, the auxiliary random variable U represents the encoded sequence for the common message and V represents the encoded sequence for both the common and confidential messages. The auxiliary random variable V_0 is both the legitimate receiver and the wiretapper's estimation of U, and the index of V_0 is related to the update information generated by the noiseless feedback. The auxiliary random*

variable V_1 is the legitimate receiver's estimation of V, and V_2 is the wiretapper's estimation of V. Both the indexes of V_1 and V_2 are with respect to the update information. The inner bound $\mathcal{C}_s^{f-new}$ is constructed by using the feedback to generate a secret key shared between the legitimate receiver and the wiretapper, and generate update information used to construct estimation of the transmitted sequences U and V. The estimation of U and V helps both the legitimate receiver and the wiretapper to improve their own received symbols Y and Z.

3. Binary Example of the BC-CM with Noiseless Feedback

Now we consider a binary case of the model of Figure 1. In this case, the channel input is X and output Y, Z takes values in $\{0,1\}$, and they satisfy

$$Y = X + Z_1, \;\; Z = X + Z_2, \tag{1}$$

where $Z_1 \sim Bern(p)$ $(p < 0.5)$ and $Z_2 \sim Bern(q)$ $(q < 0.5)$ are the channel noises for the transmitter-legitimate receiver's channel and transmitter-wiretapper's channel, respectively, and they are independent of each other and the channel input X.

Without noiseless feedback, letting $P(U=0)=\alpha$, $P(U=1)=1-\alpha$, $P(V=0)=\beta$, $P(V=1)=1-\beta$, $U+V=X$, using the fact that U is independent of V, and substituting (1) into (2), it is not difficult to calculate the secrecy capacity region $\mathcal{C}_s^b$ of the binary BC-CM, and it is given by

$$\begin{aligned}\mathcal{C}_s^b = \{&(R_0,R_1): 0 \le R_0 \le \min\{1-h(\beta\star p), 1-h(\beta\star q)\}\\ &0 \le R_1 \le h(\beta\star p) - h(p) - h(\beta\star q) + h(q)\},\end{aligned} \tag{2}$$

where $h(x) = -x\log(x) - (1-x)\log(1-x)$ and $a\star b = a+b-2ab$. Here, note that the region (2) is achieved when $\alpha = 0.5$.

With noiseless feedback, first, we compute Ahlswede and Cai's achievable secrecy rate region for this binary case. Letting $P(U=0)=\alpha$, $P(U=1)=1-\alpha$, $P(V=0)=\beta$, $P(V=1)=1-\beta$, $U+V=X$, using the fact that U is independent of V, and substituting (1) into (4), it is not difficult to calculate Ahlswede and Cai's achievable secrecy rate region $\mathcal{C}_s^{bf*}$ for this binary case, and it is given by

$$\begin{aligned}\mathcal{C}_s^{bf*} = \{&(R_0,R_1): 0 \le R_0 \le \min\{1-h(\beta\star p), 1-h(\beta\star q)\}\\ &0 \le R_1 \le \min\{h(\beta\star p) - h(p), [h(\beta\star p) - h(p) - h(\beta\star q) + h(q)]^+ + h(p)\}\},\end{aligned} \tag{3}$$

where $[x]^+ = x$ if $x \ge 0$, else $[x]^+ = 0$. Comparing (3) with (2), it is easy to see that the noiseless feedback enhances the secrecy capacity region of the binary BC-CM. Here the region (3) is achieved when $\alpha = 0.5$.

Then, it remains to compute our new achievable secrecy rate region for this binary case. Letting $V_1 = (U,V)$, $V_2 = U$, $V_0 = Z_1$, $P(U=0)=\alpha$, $P(U=1)=1-\alpha$, $P(V=0)=\beta$, $P(V=1)=1-\beta$, $U+V=X$, using the fact that U is independent of V, and substituting (1) into $\mathcal{C}_s^{f-new}$ of Theorem 1, it is not difficult to show that the achievable secrecy rate region $\mathcal{C}_s^{bf}$ of our new feedback scheme is given by

$$\begin{aligned}\mathcal{C}_s^{bf} = \{&(R_0,R_1): 0 \le R_0 \le \min\{1-h(\beta\star p) - h(p), 1-h(\beta) - h(q)\}\\ &0 \le R_1 \le \min\{h(\beta), h(\beta) - h(\beta\star q) + h(p) + h(q)\}\\ &0 \le R_1 + R_2 \le \min\{1-h(\beta) - h(p) - h(q), 1-h(\beta\star q) - h(p)\}\}.\end{aligned} \tag{4}$$

The achievability of $\mathcal{C}_s^{bf}$ can be explained by the following simple block length-(n) scheme.

- First note that in the following explanation, the channel input x^N for the i-th block ($1 \leq i \leq n$) is denoted by $\widetilde{x}_i$, and similar conventions are applied to u^N, v^N, v_0^N, v_1^N, v_2^N, y^N, z^N, z_1^N and z_2^N. For each block, the transmitted message is composed of a common message, a confidential message, a dummy message and update information.
- (*Encoding*): In the i-th block ($2 \leq i \leq n$), after the transmitter receives the feedback channel output $\widetilde{y}_{i-1}$, he generates a secret key from $\widetilde{y}_{i-1}$ and uses this key to encrypt the confidential message of the i-th block. In addition, since $\widetilde{y}_{i-1} = \widetilde{x}_{i-1} \oplus \widetilde{z}_{1,i-1}$, the transmitter also knows the legitimate receiver's channel noise $\widetilde{z}_{1,i-1}$ at the i-th block, and thus he chooses $\widetilde{v}_{0,i} = \widetilde{y}_{i-1} \oplus \widetilde{x}_{i-1} = \widetilde{z}_{1,i-1}$ as an estimation of $\widetilde{u}_{i-1}$, $\widetilde{v}_{1,i} = (\widetilde{u}_{i-1}, \widetilde{v}_{i-1})$ as the legitimate receiver's estimation of $\widetilde{x}_{i-1}$, and $\widetilde{v}_{2,i} = \widetilde{u}_{i-1}$ as the wiretapper's estimation of $\widetilde{x}_{i-1}$. Note that $\widetilde{x}_{i-1} = \widetilde{u}_{i-1} \oplus \widetilde{v}_{i-1}$ and the update information is part of the indexes of $\widetilde{v}_{0,i}$, $\widetilde{v}_{1,i}$ and $\widetilde{v}_{2,i}$.
- (*Decoding at the legitimate receiver*): The legitimate receiver does backward decoding, i.e., the decoding starts from the last block. In block n, the legitimate receiver applies Ahlswede and Cai's decoding scheme [3] to obtain his update information for block n. Then using the channel output $\widetilde{y}_n$ as side information, the legitimate receiver applies Wyner-Ziv's decoding scheme [8] to obtain $\widetilde{v}_{0,n}$ and $\widetilde{v}_{1,n}$. Since $\widetilde{v}_{0,n} = \widetilde{z}_{1,n-1}$, the legitimate receiver knows the legitimate receiver's channel noise for block $n-1$, and thus he computes $\widetilde{y}_{n-1} \oplus \widetilde{z}_{1,n-1}$ to obtain $\widetilde{x}_{n-1}$ and the corresponding transmitted message for block $n-1$. Repeating the above decoding scheme, the legitimate receiver obtains the entire transmitted messages (including both confidential and common messages) for all blocks, and since he also knows the secret keys, the real messages are decrypted by him.
- (*Decoding at the wiretapper*): The wiretapper also does backward decoding. In block n, the wiretapper receives $\widetilde{z}_n$, and he applies Ahlswede and Cai's decoding scheme [3] to obtain his update information for block n. Then using the channel output $\widetilde{z}_n$ as side information, the wiretapper applies Wyner-Ziv's decoding scheme [8] to obtain $\widetilde{v}_{0,n}$ and $\widetilde{v}_{2,n}$. Since $\widetilde{v}_{2,n} = \widetilde{u}_{n-1}$, the wiretapper knows the common message for block $n-1$. Repeating the above decoding scheme, finally, the wiretapper obtains the entire common messages for all blocks.

The following Figure 2 shows the achievable secrecy rate region $\mathcal{C}_s^{bf}$ of our new scheme, Ahlswede and Cai's achievable secrecy rate region $\mathcal{C}_s^{bf*}$ and the secrecy capacity region $\mathcal{C}_s^{b}$ of the binary BC-CM without feedback for $p = 0.05$ and $q = 0.01$, which implies that the wiretapper's channel noise is smaller than the legitimate receiver's. From Figure 2, it is easy to see that when the wiretapper's channel noise is smaller than the legitimate receiver's, the secrecy rate R_1 of the binary BC-CM without feedback is 0, which implies that perfect secrecy can not be achieved, and the secrecy rate R_1 is enhanced by using noiseless feedback. Moreover, we see that our new scheme performs better than Ahlswede and Cai's in enhancing the secrecy rate R_1, however, we should notice that the boosting of the secrecy rate R_1 is at the cost of reducing the rate R_0 of the common message.

The following Figure 3 shows the achievable secrecy rate region $\mathcal{C}_s^{bf}$ of our new scheme, Ahlswede and Cai's achievable secrecy rate region $\mathcal{C}_s^{bf*}$, and the secrecy capacity region $\mathcal{C}_s^{b}$ of the binary BC-CM without feedback for $p = 0.05$ and $q = 0.1$, which implies that the wiretapper's channel noise is larger than the legitimate receiver's. From Figure 3, it is easy to see that noiseless feedback enhances the secrecy rate of the BC-CM without feedback. However, we also should notice that the enhancement of the secrecy rate R_1 is at the cost of reducing the rate R_0 of the common message.

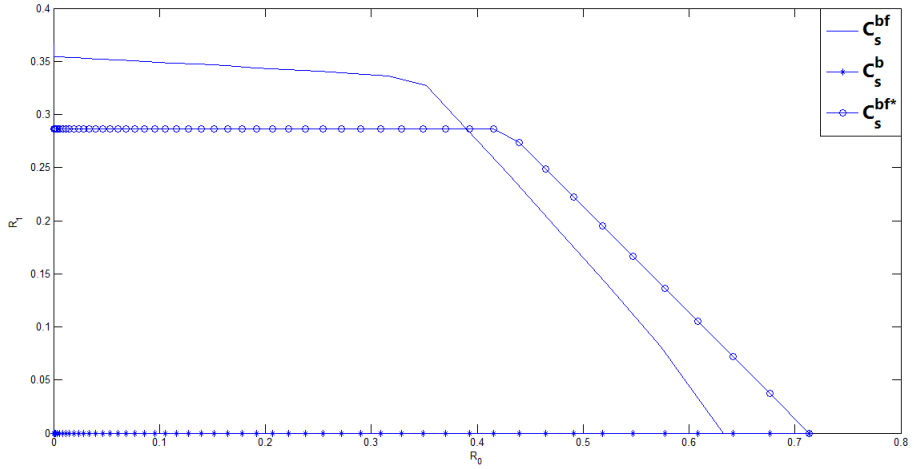

Figure 2. The comparison of our new scheme with Ahlswede-Cai's scheme and Csiszár-Körner's scheme of the BC-CM without feedback for $p = 0.05$ and $q = 0.01$.

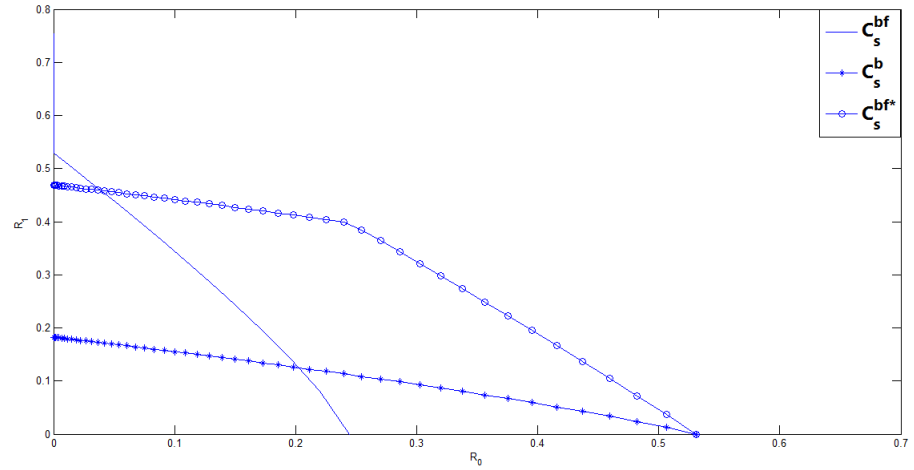

Figure 3. The comparison of our new scheme with Ahlswede-Cai's scheme and Csiszár-Körner's scheme of the BC-CM without feedback for $p = 0.05$ and $q = 0.1$.

4. Conclusions

In this paper, we propose a new coding scheme for the BC-CM with noiseless feedback. From a binary example, we show that our new feedback scheme performs better than the existing feedback scheme in enhancing the secrecy level of the BC-CM. However, we should notice that this enhancement of the secrecy level is at the cost of reducing the rate of the common message.

Acknowledgments: This work was supported by the National Natural Science Foundation of China under Grants 61671391, 61301121, 61571373 and the Open Research Fund of the State Key Laboratory of Integrated Services Networks, Xidian University (No. ISN17-13).

Author Contributions: Bin Dai, Xin Li and Zheng Ma did the theoretical work; Xin Li performed the experiments; Bin Dai and Xin Li analyzed the data; Xin Li wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
DOAJ	Directory of open access journals
TLA	Three letter acronym
LD	linear dichroism
BC-CM	broadcast channel with confidential messages

Appendix A Proof of Theorem 1

Appendix A.1 Preliminary

For a given probability $P(x)$, the identical independent distributed (i.i.d.) generated sequence x^N is called ϵ-typical if

$$|\frac{N_{x^N}(x)}{N} - P(x)| \leq \epsilon P(x),$$

where $\frac{N_{x^N}(x)}{N}$ is the frequency of symbol x appearing in the sequence x^N. The set, which is composed of all ϵ-typical x^N, is denoted by $T_\epsilon^N(P(x))$, and it is called the typical set. The following four lemmas about the typical set are extensively used in information theory.

Lemma A1. ***(Covering Lemma [9])****: Let X^N satisfy $P(X^N \in T_{\epsilon'}^N(P(x))) \to\to 1$ as $N \to \infty$. Also let M be an integer larger than 2^{Nr} for some $r \geq 0$, and let $\{Y^N(m)\}_{m=1}^M$ be a set composed of i.i.d. generated sequences Y^N (according to the probability $P(y)$) such that $\{X^N, \{Y^N(m)\}_{m=1}^M\}$ are mutually independent. Then, for any probability $P(x,y)$ with marginal probabilities $P(x)$ and $P(y)$, there exists a $\epsilon > 0$ such that*

$$\lim_{N\to\infty} P(\forall m \in \{1,2,...,M\},\ (X^N, Y^N(m)) \notin T_\epsilon^N(P(x,y))) = 0$$

if $r > I(X;Y) + \delta(\epsilon)$, where $\delta(\epsilon) \to 0$ as $\epsilon \to 0$.

Lemma A2. ***(Packing Lemma [9])****: Let X^N be an i.i.d. generated random vector with distribution $P(x)$. Also let M be an integer smaller than 2^{Nr} for some $r \geq 0$, and let $\{Y^N(m)\}_{m=1}^M$ be a set composed of i.i.d. generated sequences Y^N according to the probability $P(y)$, and each $Y^N(m)$ in the set is independent of X^N. Then for any probability $P(x,y)$ with marginal probabilities $P(x)$ and $P(y)$, there exists a $\epsilon > 0$ such that*

$$\lim_{N\to\infty} P(\exists m \in \{1,2,...,M\}\ s.t.\ (X^N, Y^N(m)) \in T_\epsilon^N(P(x,y))) = 0$$

if $r < I(X;Y) - \delta(\epsilon)$, where $\delta(\epsilon) \to 0$ as $\epsilon \to 0$.

Lemma A3. ***(Generalized Packing Lemma [9])****: For some $r_1, r_2, r_3 \geq 0$, let M_1, M_2, M_3 be integers satisfying $M_1 \leq 2^{Nr_1}$, $M_2 \leq 2^{Nr_2}$ and $M_3 \leq 2^{Nr_3}$, respectively. Also let $\{U_i^N(m)\}_{m=1}^{M_i}$ $(i = 1,2,3)$ be a set composed of i.i.d. generated sequences U_i^N (with respect to the distribution $P(u_i)$) such that $(U_1^N(m_1), U_2^N(m_2), U_3^N(m_3))$ are mutually independent for any m_1, m_2, m_3. Then for any probability $P(u_1,u_2,u_3)$ with marginal probabilities $P(u_1)$, $P(u_2)$ and $P(u_3)$, there exists a $\epsilon > 0$ such that*

$$\lim_{N\to\infty} P\left(\exists m_i \in \{1,2,...,M_i\}\ s.t.\ (U_1^N(m_1), U_2^N(m_2), U_3^N(m_3)) \in T_\epsilon^N(P(u_1,u_2,u_3))\right) = 0$$

if $r_1 + r_2 + r_3 < I(U_1;U_2) + I(U_3;U_1,U_2) - \delta(\epsilon)$, where $\delta(\epsilon) \to 0$ as $\epsilon \to 0$.

Lemma A4. ***(Balanced coloring lemma [3, p. 260])****: For any $\epsilon_1, \epsilon_2, \epsilon_3, \delta > 0$, sufficiently large N and all i.i.d. generated Y^N according to the distribution $P(y)$, there exists a γ- coloring*

$$c : T^N_{\epsilon_1}(P(y)) \to \{1, 2, .., \gamma\}$$

of $T^N_{\epsilon_1}(P(y))$ such that for all joint distribution $P(u, v, v_2, y, z)$ with marginal distribution $P(u, v, v_2, z)$ and $\frac{|T^N_{P(y|z,u,v,v_2)}(z^N, u^N, v^N, v_2^N)|}{\gamma} > 2^{N\epsilon_2}$, $(z^N, u^N, v^N, v_2^N) \in T^N_{\epsilon_3}(P(z, u, v, v_2))$,

$$|c^{-1}(k)| \leq \frac{|T^N_{P(y|z,u,v,v_2)}(z^N, u^N, v^N, v_2^N)|(1+\delta)}{\gamma}, \tag{A1}$$

for $k = 1, 2, ..., \gamma$, where c^{-1} is the inverse image of c.

Lemma A4 implies that if y^N, z^N, u^N, v^N and v_2^N are jointly typical, for given z^N, u^N, v^N and v_2^N, the number of $y^N \in T^N_{P(y|z,u,v,v_2)}(z^N, u^N, v^N, v_2^N)$ for a certain color k ($k = 1, 2, ..., \gamma$), which is denoted by $|c^{-1}(k)|$, is upper bounded by $\frac{|T^N_{P(y|z,u,v,v_2)}(z^N, u^N, v^N, v_2^N)|(1+\delta)}{\gamma}$. By using Lemma A1, it is easy to see that the typical set $T^N_{P(y|z,u,v,v_2)}(z^N, u^N, v^N, v_2^N)$ maps into at least

$$\frac{|T^N_{P(y|z,u,v,v_2)}(z^N, u^N, v^N, v_2^N)|}{\frac{|T^N_{P(y|z,u,v,v_2)}(z^N, u^N, v^N, v_2^N)|(1+\delta)}{\gamma}} = \frac{\gamma}{1+\delta} \tag{A2}$$

colors. On the other hand, the typical set $T^N_{P(y|z,u,v,v_2)}(z^N, u^N, v^N, v_2^N)$ maps into at most γ colors.

Appendix A.2 Code Construction

Definitions:

- Transmission takes place over n blocks, and each block is of length N. Define the confidential message W_1 by $W_1 = (W_{1,1}, ..., W_{1,n})$, where $W_{1,i}$ ($1 \leq i \leq n$) is for block i and takes values in $\{1, 2, ..., 2^{NR_1}\}$. Further divide $W_{1,i}$ into $W_{1,i} = (W_{1,1,i}, W_{1,2,i})$, where $W_{1,j,i}$ ($j = 1, 2$) takes values in $\{1, 2, ..., 2^{NR_{1,j}}\}$, and $R_{1,1} + R_{1,2} = R_1$.
- Define the common message W_0 by $W_0 = (W_{0,1}, ..., W_{0,n})$, where $W_{0,i}$ ($1 \leq i \leq n$) is for block i and takes values in $\{1, 2, ..., 2^{NR_0}\}$.
- Let $W^{'}$ be a randomly generated dummy message transmitted over all blocks, and it is denoted by $W^{'} = (W^{'}_1, ..., W^{'}_n)$, where $W^{'}_i$ ($1 \leq i \leq n$) is for block i and it takes values in $\{1, 2, ..., 2^{NR^{'}}\}$.
- Let W^*_0 and W^*_1 be update information transmitted over all blocks, and they are respectively denoted by $W^*_0 = (W^*_{0,1}, ..., W^*_{0,n})$ and $W^*_1 = (W^*_{1,1}, ..., W^*_{1,n})$, where $W^*_{0,i}$ and $W^*_{1,i}$ ($1 \leq i \leq n$) are for block i and take values in $\{1, 2, ..., 2^{N\tilde{R}_0}\}$ and $\{1, 2, ..., 2^{N\tilde{R}_1}\}$, respectively. Further divide $W^*_{0,i}$ into $W^*_{0,i} = (W^*_{0,0,i}, W^*_{0,1,i}, W^*_{0,2,i})$, where $W^*_{0,j,i}$ ($j = 0, 1, 2$) takes values in $\{1, 2, ..., 2^{N\tilde{R}_{0,j}}\}$, and $\tilde{R}_{0,0} + \tilde{R}_{0,1} + \tilde{R}_{0,2} = \tilde{R}_0$. Moreover, further divide $W^*_{1,i}$ into $W^*_{1,i} = (W^*_{1,0,i}, W^*_{1,1,i})$, where $W^*_{1,j,i}$ ($j = 0, 1$) takes values in $\{1, 2, ..., 2^{N\tilde{R}_{1,j}}\}$, and $\tilde{R}_{1,0} + \tilde{R}_{1,1} = \tilde{R}_1$.
- Let $\widetilde{X}_i$, $\widetilde{Y}_i$, $\widetilde{Z}_i$, $\widetilde{U}_i$, $\widetilde{V}_i$, $\widetilde{V}_{0,i}$, $\widetilde{V}_{1,i}$ and $\widetilde{V}_{2,i}$ be the random vectors for block i ($1 \leq i \leq n$). Define $X^n = (\widetilde{X}_1, ..., \widetilde{X}_n)$, and similar convention is applied to Y^n, Z^n, U^n, V^n, V^n_0, V^n_1 and V^n_2. The specific values of the above random vectors are denoted by lower case letters.

Code construction:

- In each block i ($1 \leq i \leq n$), randomly produce $2^{N(R_0+\tilde{R}_0)}$ i.i.d. sequences $\widetilde{u}_i$ according to the probability $P(u)$, and index them as $\widetilde{u}_i(w_{0,i}, w^*_{0,0,i}, w^*_{0,1,i}, w^*_{0,2,i})$, where $w_{0,i} \in \{1, 2, ..., 2^{NR_0}\}$, $w^*_{0,0,i} \in \{1, 2, ..., 2^{N\tilde{R}_{0,0}}\}$, $w^*_{0,1,i} \in \{1, 2, ..., 2^{N\tilde{R}_{0,1}}\}$ and $w^*_{0,2,i} \in \{1, 2, ..., 2^{N\tilde{R}_{0,2}}\}$. Here note that $\tilde{R}_{0,0} + \tilde{R}_{0,1} + \tilde{R}_{0,2} = \tilde{R}_0$.

- For a given $\widetilde{u}_i(w_{0,i}, w^*_{0,0,i}, w^*_{0,1,i}, w^*_{0,2,i})$, randomly produce $2^{N(R_1+R^{'}+\tilde{R}_1)}$ i.i.d. sequences $\widetilde{v}_i$ according to the conditional probability $P(v|u)$, and index them as $\widetilde{v}_i(w_{1,1,i}, w_{1,2,i}, w^{'}_i, w^*_{1,0,i}, w^*_{1,1,i})$, where $w_{1,1,i} \in \{1, 2, ..., 2^{NR_{1,1}}\}$, $w_{1,2,i} \in \{1, 2, ..., 2^{NR_{1,2}}\}$, $w^{'}_i \in \{1, 2, ..., 2^{NR^{'}}\}$, $w^*_{1,0,i} \in \{1, 2, ..., 2^{N\tilde{R}_{1,0}}\}$ and $w^*_{1,1,i} \in \{1, 2, ..., 2^{N\tilde{R}_{1,1}}\}$. Here note that $R_{1,1} + R_{1,2} = R_1$ and $\tilde{R}_{1,0} + \tilde{R}_{1,1} = \tilde{R}_1$.
- The sequence $\widetilde{x}_i$ is i.i.d. produced according to a new discrete memoryless channel (DMC) with transition probability $P(x|u, v)$. The inputs and output of this new DMC are $\widetilde{u}_i$, $\widetilde{v}_i$ and $\widetilde{x}_i$, respectively.
- In each block i ($1 \leq i \leq n$), generate $\widetilde{v}_{0,i}$ in two ways: the first way is to produce $2^{N(\tilde{R}_{0,0}+\tilde{R}^{'}_0)}$ i.i.d. sequences $\widetilde{v}_{0,i}$ according to the probability $P(v_0|u, v, y)$, and index them as $\widetilde{v}_{0,i}(1; w^*_{0,0,i}, w^*_{1,0,i}, l_{1,0,i})$, where 1 represents the first way to define $\widetilde{v}_{0,i}$, $w^*_{0,0,i} \in \{1, 2, ..., 2^{N\tilde{R}_{0,0}}\}$, $w^*_{1,0,i} \in \{1, 2, ..., 2^{N\tilde{R}_{1,0}}\}$ and $l_{1,0,i} \in \{1, 2, ..., 2^{N(\tilde{R}^{'}_0-\tilde{R}_{1,0})}\}$; the second way is to produce $2^{N(\tilde{R}_{0,0}+\tilde{R}^{'}_0)}$ i.i.d. sequences $\widetilde{v}_{0,i}$ according to the probability $P(v_0|u, v, y)$, and index them as $\widetilde{v}_{0,i}(2; w^*_{0,0,i}, l_{2,0,i})$, where 2 represents the second way to define $\widetilde{v}_{0,i}$, $w^*_{0,0,i} \in \{1, 2, ..., 2^{N\tilde{R}_{0,0}}\}$, and $l_{2,0,i} \in \{1, 2, ..., 2^{N\tilde{R}^{'}_0}\}$.
- In each block i ($1 \leq i \leq n$), produce $2^{N(\tilde{R}_{0,1}+\tilde{R}_{1,1}+\tilde{R}^{'}_1)}$ i.i.d. sequences $\widetilde{v}_{1,i}$ according to the probability $P(v_1|u, v, y)$, and index them as $\widetilde{v}_{1,i}(w^*_{0,1,i}, w^*_{1,1,i}, l_{1,i})$, where $w^*_{0,1,i} \in \{1, 2, ..., 2^{N\tilde{R}_{0,1}}\}$, $w^*_{1,1,i} \in \{1, 2, ..., 2^{N\tilde{R}_{1,1}}\}$ and $l_{1,i} \in \{1, 2, ..., 2^{N\tilde{R}^{'}_1}\}$.
- In each block i ($1 \leq i \leq n$), produce $2^{N(\tilde{R}_{0,2}+\tilde{R}^{'}_2)}$ i.i.d. sequences $\widetilde{v}_{2,i}$ according to the probability $P(v_2|u, v, y)$, and index them as $\widetilde{v}_{2,i}(w^*_{0,2,i}, l_{2,i})$, where $w^*_{0,2,i} \in \{1, 2, ..., 2^{N\tilde{R}_{0,2}}\}$ and $l_{2,i} \in \{1, 2, ..., 2^{N\tilde{R}^{'}_2}\}$.

Encoding scheme:

- In block 1, the transmitter chooses $\widetilde{u}_1(w_{0,1}, 1, 1, 1)$ and $\widetilde{v}_1(w_{1,1,1}, w_{1,2,1} = 1, w^{'}_1, 1, 1)$ to transmit.
- In block i ($2 \leq i \leq n-1$), the transmitter receives the feedback $\widetilde{y}_{i-1}$, and he tries to select a pair of sequences $(\widetilde{v}_{0,i-1}, \widetilde{v}_{1,i-1})$ such that $(\widetilde{v}_{0,i-1}(1; w^*_{0,0,i-1}, w^*_{1,0,i-1}, l_{1,0,i-1}), \widetilde{v}_{1,i-1}(w^*_{0,1,i-1}, w^*_{1,1,i-1}, l_{1,i-1}), \widetilde{u}_{i-1}, \widetilde{v}_{i-1}, \widetilde{y}_{i-1})$ are jointly typical sequences. If there are more than one pair $(\widetilde{v}_{0,i-1}, \widetilde{v}_{1,i-1})$, randomly choose one; if there is no such pair, an error is declared. Based on Lemma A1, it is easy to see that the error probability goes to 0 if

$$\tilde{R}_{0,0} + \tilde{R}^{'}_0 \geq I(V_0; U, V, Y), \tag{A3}$$

$$\tilde{R}_{0,1} + \tilde{R}_{1,1} + \tilde{R}^{'}_1 \geq I(V_1; U, V, Y, V_0). \tag{A4}$$

Moreover, the transmitter also tries to select a pair of sequences $(\widetilde{v}_{0,i-1}, \widetilde{v}_{2,i-1})$ such that $(\widetilde{v}_{0,i-1}(2; w^*_{0,0,i-1}, l_{2,0,i-1}), \widetilde{v}_{2,i-1}(w^*_{0,2,i-1}, l_{2,i-1}), \widetilde{u}_{i-1}, \widetilde{v}_{i-1}, \widetilde{y}_{i-1})$ are jointly typical sequences. If there are more than one pair $(\widetilde{v}_{0,i-1}, \widetilde{v}_{2,i-1})$, randomly choose one; if there is no such pair, an error is declared. Based on Lemma A1, it is easy to see that the error probability goes to 0 if (A3) and

$$\tilde{R}_{0,2} + \tilde{R}^{'}_2 \geq I(V_2; U, V, Y, V_0) \tag{A5}$$

hold. Once the transmitter selects such pairs $(\widetilde{v}_{0,i-1}, \widetilde{v}_{1,i-1})$ and $(\widetilde{v}_{0,i-1}, \widetilde{v}_{2,i-1})$, he chooses $\widetilde{u}_i(w_{0,i}, w^*_{0,0,i-1}, w^*_{0,1,i-1}, w^*_{0,2,i-1})$ to transmit.

Before choosing the transmitted codeword $\widetilde{v}_i$, produce a mapping $g_i : \widetilde{y}_{i-1} \to \{1, 2, ..., 2^{NR_{1,2}}\}$. Furthermore, we define $K_i = g_i(\widetilde{Y}_{i-1})$ as a random variable uniformly distributed over $\{1, 2, ..., 2^{NR_{1,2}}\}$, and it is independent of all the random vectors and messages of block i. Here note that K_i is the secret key known by the transmitter and the legitimate receiver, and $k_i = g_i(\widetilde{y}_{i-1}) \in \{1, 2, ..., 2^{NR_2}\}$ is a specific value of K_i. Reveal the mapping g_i to the transmitters, legitimate receiver and the wiretapper. Once the transmitter finds a pair $(\widetilde{v}_{0,i-1}, \widetilde{v}_{1,i-1})$ such that $(\widetilde{v}_{0,i-1}(1; w^*_{0,0,i-1}, w^*_{1,0,i-1}, l_{1,0,i-1}), \widetilde{v}_{1,i-1}(w^*_{0,1,i-1}, w^*_{1,1,i-1}, l_{1,i-1}), \widetilde{u}_{i-1}, \widetilde{v}_{i-1}, \widetilde{y}_{i-1})$ are jointly typical sequences, and finds a pair $(\widetilde{v}_{0,i-1}, \widetilde{v}_{2,i-1})$ such that $(\widetilde{v}_{0,i-1}(2; w^*_{0,0,i-1}, l_{2,0,i-1}), \widetilde{v}_{2,i-1}(w^*_{0,2,i-1}, l_{2,i-1}), \widetilde{u}_{i-1}, \widetilde{v}_{i-1}, \widetilde{y}_{i-1})$ are jointly typical sequences, he chooses $\widetilde{v}_i(w_{1,1,i}, w_{1,2,i} \oplus k_i, w_i^{'}, w^*_{1,0,i-1}, w^*_{1,1,i-1})$ to transmit.

- In block n, the transmitter receives $\widetilde{y}_{n-1}$, and he finds a pair $(\widetilde{v}_{0,n-1}, \widetilde{v}_{1,n-1})$ such that $(\widetilde{v}_{0,n-1}(1; w^*_{0,0,n-1}, w^*_{1,0,n-1}, l_{1,0,n-1}), \widetilde{v}_{1,n-1}(w^*_{0,1,n-1}, w^*_{1,1,n-1}, l_{1,n-1}), \widetilde{u}_{n-1}, \widetilde{v}_{n-1}, \widetilde{y}_{n-1})$ are jointly typical sequences. Moreover, he also finds a pair $(\widetilde{v}_{0,n-1}, \widetilde{v}_{2,n-1})$ such that $(\widetilde{v}_{0,n-1}(2; w^*_{0,0,n-1}, l_{2,0,n-1}), \widetilde{v}_{2,n-1}(w^*_{0,2,n-1}, l_{2,n-1}), \widetilde{u}_{n-1}, \widetilde{v}_{n-1}, \widetilde{y}_{n-1})$ are jointly typical sequences. Then he chooses $\widetilde{u}_n(1, w^*_{0,0,n-1}, w^*_{0,1,n-1}, w^*_{0,2,n-1})$ and $\widetilde{v}_n(1, 1, 1, w^*_{1,0,n-1}, w^*_{1,1,n-1})$ to transmit.

Decoding scheme for the legitimate receiver: The legitimate receiver does backward decoding after the transmission of all n blocks is finished. For block n, first, he tries to select a unique $\widetilde{u}_n$ such that $(\widetilde{u}_n, \widetilde{y}_n)$ are jointly typical. If there is no $\widetilde{u}_n$ or multiple ones exist, an decoding error is declared. Using Lemma A2, the error probability goes to 0 if

$$\tilde{R}_{0,0} + \tilde{R}_{0,1} + \tilde{R}_{0,2} \leq I(U; Y). \tag{A6}$$

Then, he tries to select a unique $\widetilde{v}_n$ such that $(\widetilde{u}_n, \widetilde{v}_n, \widetilde{y}_n)$ are jointly typical. If there is no $\widetilde{v}_n$ or multiple ones exist, an decoding error is declared. Using Lemma A2, the error probability goes to 0 if

$$\tilde{R}_{1,0} + \tilde{R}_{1,1} \leq I(V; Y|U). \tag{A7}$$

When $\widetilde{u}_n$ and $\widetilde{v}_n$ are successfully decoded, the legitimate receiver extracts $w^*_{0,0,n-1}$, $w^*_{0,1,n-1}$, $w^*_{0,2,n-1}$, $w^*_{1,0,n-1}$, $w^*_{1,1,n-1}$ from them. Then using Wyner-Ziv's decoding scheme [8] for the source coding with side information, the legitimate receiver tries to find unique $\widetilde{v}_{0,n-1}$ and $\widetilde{v}_{1,n-1}$ such that given $w^*_{0,0,n-1}$, $w^*_{0,1,n-1}$, $w^*_{0,2,n-1}$, $w^*_{1,0,n-1}$ and $w^*_{1,1,n-1}$, $(\widetilde{v}_{0,n-1}, \widetilde{v}_{1,n-1}, \widetilde{y}_{n-1})$ are jointly typical sequences. If there is no $\widetilde{v}_{1,n-1}$ or multiple ones exist, an decoding error is declared. Based on Lemma A2 and Lemma A3, the error probability goes to 0 if

$$\tilde{R}_1^{'} \leq I(V_1; V_0, Y), \tag{A8}$$

$$\tilde{R}_1^{'} + \tilde{R}_0^{'} - \tilde{R}_{1,0} \leq I(V_0; Y) + I(V_1; V_0, Y). \tag{A9}$$

For block $n-1$, after $\widetilde{v}_{1,n-1}$ is successfully decoded, the legitimate receiver tries to select a unique $\widetilde{u}_{n-1}$ such that $(\widetilde{u}_{n-1}, \widetilde{y}_{n-1}, \widetilde{v}_{1,n-1})$ are jointly typical. Based on Lemma A2, the error probability goes to 0 if

$$R_0 + \tilde{R}_0 \leq I(U; Y, V_1). \tag{A10}$$

Then he tries to select a unique $\widetilde{v}_{n-1}$ such that $(\widetilde{u}_{n-1}, \widetilde{v}_{n-1}, \widetilde{y}_{n-1}, \widetilde{v}_{1,n-1})$ are jointly typical. If there is no $\widetilde{v}_{n-1}$ or multiple ones exist, an decoding error is declared. Using Lemma A2, the error probability goes to 0 if

$$R_{1,1} + R_{1,2} + R^{'} + \tilde{R}_{1,0} + \tilde{R}_{1,1} \leq I(V; Y, V_1|U). \tag{A11}$$

When $\widetilde{u}_{n-1}$ and $\widetilde{v}_{n-1}$ are successfully decoded, the legitimate receiver extracts $w_{0,n-1}$, $w_{1,1,n-1}$, $w_{1,2,n-1} \oplus k_{n-1}$, w'_{n-1}, $w^*_{0,0,n-2}$, $w^*_{0,1,n-2}$, $w^*_{0,2,n-2}$, $w^*_{1,0,n-2}$ and $w^*_{1,1,n-2}$ from it. Since the legitimate receiver knows the key $k_{n-1} = g_{n-1}(\widetilde{y}_{n-2})$, the transmitted messages $w_{0,n-1}$, $w_{1,1,n-1}$ and $w_{1,2,n-1}$ for block $n-1$ are obtained. Repeat the above decoding scheme, the entire transmitted messages for all blocks are obtained by the legitimate receiver.

Decoding scheme for the wiretapper: The wiretapper also does backward decoding after the transmission of all n blocks is finished. For block n, first, he tries to select a unique $\widetilde{u}_n$ such that $(\widetilde{u}_n, \widetilde{z}_n)$ are jointly typical. If there is no $\widetilde{u}_n$ or multiple ones exist, an decoding error is declared. Using Lemma A2, the error probability goes to 0 if

$$\tilde{R}_{0,0} + \tilde{R}_{0,1} + \tilde{R}_{0,2} \leq I(U;Z). \tag{A12}$$

When $\widetilde{u}_n$ is successfully decoded, the wiretapper extracts $w^*_{0,0,n-1}$, $w^*_{0,1,n-1}$ and $w^*_{0,2,n-1}$ from them. Then using Wyner-Ziv's decoding scheme [8] for the source coding with side information, the wiretapper tries to find unique $\widetilde{v}_{0,n-1}$ and $\widetilde{v}_{2,n-1}$ such that given $w^*_{0,0,n-1}$, $w^*_{0,1,n-1}$ and $w^*_{0,2,n-1}$, $(\widetilde{v}_{0,n-1}, \widetilde{v}_{2,n-1}, \widetilde{z}_{n-1})$ are jointly typical sequences. If there is no $\widetilde{v}_{2,n-1}$ or multiple ones exist, an decoding error is declared. Based on Lemma A2 and Lemma A3, the error probability goes to 0 if

$$\tilde{R}'_2 \leq I(V_2; V_0, Z), \tag{A13}$$

$$\tilde{R}'_2 + \tilde{R}'_0 \leq I(V_0; Z) + I(V_2; V_0, Z). \tag{A14}$$

For block $n-1$, after $\widetilde{v}_{2,n-1}$ is successfully decoded, the wiretapper tries to select a unique $\widetilde{u}_{n-1}$ such that $(\widetilde{u}_{n-1}, \widetilde{z}_{n-1}, \widetilde{v}_{2,n-1})$ are jointly typical. Based on Lemma A2, the error probability goes to 0 if

$$R_0 + \tilde{R}_0 \leq I(U; Z, V_2). \tag{A15}$$

When $\widetilde{u}_{n-1}$ is successfully decoded, the legitimate receiver extracts $w_{0,n-1}$, $w^*_{0,0,n-2}$, $w^*_{0,1,n-2}$ and $w^*_{0,2,n-2}$ from it. Repeat the above decoding scheme, the entire common messages for all blocks are obtained by the wiretapper.

Appendix A.3 Equivocation Analysis

For all blocks, the equivocation Δ is bounded by

$$\begin{aligned} \Delta &= \frac{1}{nN} H(W_0, W_1 | Z^n) \geq \frac{1}{nN} H(W_1 | Z^n, W_0) \\ &\overset{(a)}{=} \frac{1}{nN} (H(W_{11} | Z^n, W_0) + H(W_{12} | Z^n, W_0, W_{11})), \end{aligned} \tag{A16}$$

where (a) is from the definitions $W_{11} = (W_{1,1,1}, ..., W_{1,1,n})$ and $W_{12} = (W_{1,2,1}, ..., W_{1,2,n})$.

The first part $H(W_{11}|Z^n, W_0)$ of (A16) can be lower bounded by

$$\begin{aligned}
&H(W_{11}|Z^n, W_0) \geq H(W_{11}|Z^n, W_0, U^n, V_2^n) \\
&\overset{(b)}{=} H(W_{11}|Z^n, U^n, V_2^n) \\
&= H(W_{11}, Z^n, U^n, V_2^n) - H(Z^n, U^n, V_2^n) \\
&= H(W_{11}, Z^n, U^n, V_2^n, V^n) - H(V^n|W_{11}, Z^n, U^n, V_2^n) - H(Z^n, U^n, V_2^n) \\
&\overset{(c)}{=} H(Z^n, V_2^n|U^n, V^n) + H(U^n, V^n) - H(V^n|W_{11}, Z^n, U^n, V_2^n) - H(Z^n, V_2^n|U^n) - H(U^n) \\
&= H(V^n|U^n) - H(V^n|W_{11}, Z^n, U^n, V_2^n) - I(Z^n, V_2^n; V^n|U^n) \\
&\overset{(d)}{=} (n-1)NR_{1,1} + (n-2)NR_{1,2} + (n-1)NR^{'} + (n-1)N(\tilde{R}_{1,0} + \tilde{R}_{1,1}) \\
&-H(V^n|W_{11}, Z^n, U^n, V_2^n) - I(Z^n, V_2^n; V^n|U^n) \\
&\overset{(e)}{\geq} (n-1)NR_{1,1} + (n-2)NR_{1,2} + (n-1)NR^{'} + (n-1)N(\tilde{R}_{1,0} + \tilde{R}_{1,1}) \\
&-nNI(V; Z, V_2|U) - H(V^n|W_{11}, Z^n, U^n, V_2^n) \\
&\overset{(f)}{\geq} (n-1)NR_{1,1} + (n-2)NR_{1,2} + (n-1)NR^{'} + (n-1)N(\tilde{R}_{1,0} + \tilde{R}_{1,1}) \\
&-nNII(V; Z, V_2|U) - nN\epsilon^{'},
\end{aligned} \tag{A17}$$

where (b) is from $H(W_0|U^n) = 0$, (c) is from $H(W_{11}|V^n) = 0$, (d) is from the construction of U^n and V^n, (e) is from the fact that the channel is memoryless, and (f) is from the fact that given w_{11}, u^n, v_2^n and z^n, the wiretapper tries to select unique v^n such that (v^n, z^n, v_2^n, u^n) are jointly typical, and based on Lemma A2, the wiretapper's decoding error probability tends to 0 if

$$R_{1,2} + R^{'} + \tilde{R}_{1,0} + \tilde{R}_{1,1} \leq I(V; Z, V_2|U), \tag{A18}$$

then using Fano's inequality, we have $\frac{1}{nN} H(V^n|W_{11}, Z^n, U^n, V_2^n) \leq \epsilon^{'}$, where $\epsilon^{'} \to 0$ as $n, N \to \infty$.

Moreover, the second part $H(W_{12}|Z^n, W_0, W_{11})$ of (A16) can be lower bounded by

$$\begin{aligned}
&H(W_{12}|Z^n, W_0, W_{11}) \\
&\geq \sum_{i=2}^{n-1} H(W_{1,2,i}|Z^n, W_0, W_{11}, W_{1,2,1} = 1, ..., W_{1,2,i-1}, W_{1,2,i} \oplus K_i) \\
&\overset{(g)}{=} \sum_{i=2}^{n-1} H(W_{1,2,i}|\widetilde{Z}_{i-1}, W_{1,2,i} \oplus K_i) \\
&\geq \sum_{i=2}^{n-1} H(W_{1,2,i}|\widetilde{Z}_{i-1}, \widetilde{U}_{i-1}, W_{1,2,i} \oplus K_i) \\
&= \sum_{i=2}^{n-1} H(K_i|\widetilde{Z}_{i-1}, \widetilde{U}_{i-1}, W_{1,2,i} \oplus K_i) \\
&\overset{(f)}{=} \sum_{i=2}^{n-1} H(K_i|\widetilde{Z}_{i-1}, \widetilde{U}_{i-1}) \\
&\geq \sum_{i=2}^{n-1} H(K_i|\widetilde{Z}_{i-1}, \widetilde{U}_{i-1}, \widetilde{V}_{i-1}, \widetilde{V}_{2,i-1}),
\end{aligned} \tag{A19}$$

where (e) is from the Markov chain $W_{1,2,i} \to (\widetilde{Z}_{i-1}, W_{1,2,i} \oplus K_i) \to (W_0, W_{11}, W_{1,2,1}, ..., W_{1,2,i-1}, \widetilde{Z}_1, ..., \widetilde{Z}_{i-2}, \widetilde{Z}_i, ..., \widetilde{Z}_n)$, (f) is from $K_i \to (\widetilde{Z}_{i-1}, \widetilde{U}_{i-1}) \to W_{1,2,i} \oplus K_i$. Now it remains to bound $H(K_i|\widetilde{Z}_{i-1}, \widetilde{U}_{i-1}, \widetilde{V}_{i-1}, \widetilde{V}_{2,i-1})$ in (A19), see the followings.

From Lemma A4 and (A2), we know that the typical set $T^N_{P(y|z,u,v,v_2)}(z^N, u^N, v^N, v_2^N)$ maps into at least $\frac{\gamma}{1+\delta}$ colors. Choosing $\gamma = |T^N_{P(y|z,u,v,v_2)}(z^N, u^N, v^N, v_2^N)|$ and notice that

$$\begin{aligned} &|T^N_{P(y|z,u,v,v_2)}(z^N, u^N, v^N, v_2^N)| \\ &\geq (1-\epsilon_1)2^{N(1-\epsilon_2)H(Y|U,V,V_2,Z)}, \end{aligned} \tag{A20}$$

where $\epsilon_1, \epsilon_2 \to 0$ as $N \to \infty$, thus we can conclude that

$$\begin{aligned} H(K_i|\widetilde{Z}_{i-1}, \widetilde{U}_{i-1}, \widetilde{V}_{i-1}, \widetilde{V}_{2,i-1}) &\geq \log \frac{\gamma}{1+\delta} \\ &\geq \log \frac{1-\epsilon_1}{1+\delta} + N(1-\epsilon_2)H(Y|U,V,V_2,Z). \end{aligned} \tag{A21}$$

Substituting (A21) into (A19), we have

$$H(W_{12}|Z^n, W_0, W_{11}) \geq (n-2)\log\frac{1-\epsilon_1}{1+\delta} + (n-2)N(1-\epsilon_2)H(Y|U,V,V_2,Z). \tag{A22}$$

Finally, substituting (A17) and (A22) into (A16), we have

$$\begin{aligned} &\Delta \geq \frac{n-1}{n}(R_{1,1} + R^{'} + \tilde{R}_{1,0} + \tilde{R}_{1,1}) + \frac{n-2}{n}R_{1,2} - I(V;Z,V_2|U) - \epsilon^{'} \\ &+\frac{n-2}{nN}\log\frac{1-\epsilon_1}{1+\delta} + \frac{n-2}{n}(1-\epsilon_2)H(Y|U,V,V_2,Z). \end{aligned} \tag{A23}$$

The bound (A23) implies that if

$$R^{'} + \tilde{R}_{1,0} + \tilde{R}_{1,1} \geq I(V;Z,V_2|U) - H(Y|U,V,V_2,Z) \tag{A24}$$

we can prove that $\Delta \geq R_{1,1} + R_{1,2} - \epsilon$ by choosing sufficiently large n and N.

The achievable secrecy rate region can be obtained from (A3), (A4), (A5), (A6), (A7), (A8), (A9), (A10), (A11), (A12), (A13), (A14), (A15), (A18) and (A24). To be specific, first, using $\tilde{R}_0 = \tilde{R}_{0,0} + \tilde{R}_{0,1} + \tilde{R}_{0,2}$, $\tilde{R}_1 = \tilde{R}_{1,0} + \tilde{R}_{1,1}$, the Markov chain $(V_0, V_1, V_2) \to (U, V, Y) \to (Y, Z)$, and applying Fourier-Motzkin elimination to eliminate $\tilde{R}_{0,0}$, $\tilde{R}_{0,1}$, $\tilde{R}_{0,2}$, $\tilde{R}_{1,0}$, $\tilde{R}_{1,1}$, $\tilde{R}_0^{'}$, $\tilde{R}_1^{'}$ and $\tilde{R}_2^{'}$ from (A3), (A4), (A5), (A6), (A7), (A8), (A9), (A12), (A13) and (A14), we have

$$\tilde{R}_0 \geq I(U,V,Y;V_0,V_2|Z), \tag{A25}$$

$$\begin{aligned} \tilde{R}_0 + \tilde{R}_1 &\geq I(U,V,Y;V_1|V_0,Y) + I(U,V,Y;V_2|V_0,Z) \\ &+ \max\{I(V_0;U,V,Y|Y), I(V_0;U,V,Y|Z)\}. \end{aligned} \tag{A26}$$

Then, using $R_1 = R_{1,1} + R_{1,2}$, $\tilde{R}_1 = \tilde{R}_{1,0} + \tilde{R}_{1,1}$, and applying Fourier-Motzkin elimination to eliminate $R_{1,1}$, $R_{1,2}$, $\tilde{R}_{1,0}$, $\tilde{R}_{1,1}$, $R^{'}$ from (A10), (A11), (A15), (A18), (A24), (A25) and (A26), the achievable secrecy rate region C_s^{f-new} in Theorem 1 is obtained. The proof of Theorem 1 is completed.

References

1. Wyner, A.D. The wire-tap channel. *Bell Syst. Tech. J.* **1975**, *54*, 1355–1387.
2. Csiszár, I.; Körner, J. Broadcast channels with confidential messages. *IEEE Trans. Inf. Theory* **1978**, *24*, 339–348.
3. Ahlswede, R.; Cai, N. Transmission, identification and common randomness capacities for wire-tap channels with secure feedback from the decoder. *Gen. Theory Inf. Trans. Comb.* **2006**, 258–275.
4. Ardestanizadeh, E.; Franceschetti, M.; Javidi, T.; Kim, Y. Wiretap channel with secure rate-limited feedback. *IEEE Trans. Inf. Theory* **2009**, *55*, 5353–5361.

5. Lai, L.; El Gamal, H.; Poor, V. The wiretap channel with feedback: encryption over the channel. *IEEE Trans. Inf. Theory* **2008**, *54*, 5059–5067.
6. Yin, X.; Xue, Z.; Dai, B. Capacity-equivocation regions of the DMBCs with noiseless feedback. *Math. Probl. Eng.* **2013**, *2013*, 102069.
7. Dai, B.; Han Vinck, A.J.; Luo, Y.; Zhuang, Z. Capacity region of non-degraded wiretap channel with noiseless feedback. In Proceedings of 2012 IEEE International Symposium on Information Theory (ISIT), Cambridge, MA, USA, 1–6 July 2012.
8. Wyner, A.; Ziv, J. The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. Inf. Theory* **1976**, *22*, 1–10.
9. El Gamal, A.; Kim, Y.H. Information measures and typicality. In *Network Information Theory*; Cambridge University Press: Cambridge, UK, 2011; pp. 17–37.