New Directions in the Theory of Identification via Channels

Rudolf Ahlswede and Zhen Zhang, *Senior Member, IEEE*

***Abstract*—We study two problems in the theory of identification via channels. The first problem concerns the identification via channels with noisy feedback. Whereas for Shannon's transmission problem the capacity of a discrete memoryless channel does not change with feedback, we know that the identification capacity is affected by feedback. We study its dependence on the feedback channel. We prove both, a direct and a converse coding theorem. Although a gap exists between the upper and lower bounds provided by these two theorems, the known result for channels without feedback and the known result for channels with complete feedback, are both special cases of these two new theorems, because in these cases the bounds coincide. The second problem is the identification via wiretap channels. A secrecy identification capacity is defined for the wiretap channel. A "Dichotomy Theorem" is proved which says here that the second-order secrecy identification capacity is the same as Shannon's capacity for the main channel as long as the secrecy transmission capacity of the wiretap channel is not zero, and zero otherwise. Equivalently, we can say that the identification capacity is not lowered by the presence of a wiretapper as long as 1 bit can be transmitted (or identified) correctly with arbitrarily small error probability. This is in strong contrast to the case of transmission.**

***Index Terms*— Cryptography, identification, channel capacity, wiretap channel, noisy feedback.**

I. Introduction

The MODEL of identification via channels was introduced by Ahlswede and Dueck in [3]. Since then, several articles (c.f. [4]–[7]) on this subject have appeared and there are still many interesting and important open problems in this fertile research area. In this paper we introduce and study two new problems. The first concerns identification via channels with *noisy* feedback, which is a model that unifies both the case of channels without feedback and the case of channels with complete (or noiseless) feedback. The identification problems for these two cases were studied in [3] and [4], which contain the most basic results in this area.

A communication channel with noisy feedback is defined by a quadruple

$$\{\mathcal{X},\, W,\, \mathcal{Y},\, \mathcal{Z}\} \tag{1}$$

where $\mathcal{X}$ is the input alphabet, $\mathcal{Y}$ is the output alphabet, $\mathcal{Z}$ is the output alphabet for the feedback, and $W = \{W(y,\, z \mid x)\colon x \in \mathcal{X},\, y \in \mathcal{Y},\, z \in \mathcal{Z}\}$ is a stochastic matrix which gives the conditional probability of the output letters y and z when the input letter is x. The transmission probability for n-sequences $x^n = (x_1, \cdots, x_n) \in \mathcal{X}^n$, $y^n = (y_1, \cdots, y_n) \in \mathcal{Y}^n$, $z^n = (z_1, \cdots, z_n) \in \mathcal{Z}^n$ is given by

$$W^n(y^n,\, z^n \mid x^n) = \prod_{t=1}^{n} W(y_t,\, z_t \mid x_t) \tag{2}$$

for $n = 1, 2, 3, \cdots$, that is, the channel is assumed to be memoryless.

To define identification feedback codes (IDF) in the sense of [4] for this channel we let $\mathcal{F}_n$ be the set of all possible vector valued functions

$$f = [f^1, \cdots, f^n] \tag{3}$$

where for $t \in \{2, \cdots, n\}$, f^t is defined on $\mathcal{Z}^{t-1}$ and takes values in $\mathcal{X}$. f^1 is an element of $\mathcal{X}$. It is understood that, when f is used for the transmission over the channel, after the feedback signals $z_1, z_2, \cdots, z_{t-1}$ have been made known to the sender by the feedback channel, the sender transmits $f^t(z_1, \cdots, z_{t-1})$. When $t = 1$, the sender transmits f^1. The joint distribution of the output random variables $Y_1, \cdots, Y_n$ and the feedback random variables $Z_1, \cdots, Z_n$ is determined by the function f used and by W as follows. For $y^n \in \mathcal{Y}^n$, $z^n \in \mathcal{Z}^n$

$$\begin{aligned}\Pr(Y^n = y^n,\, Z^n = z^n \mid f) &= W^n(y^n,\, z^n \mid f)\\ &= \prod_{t=1}^{n} W(y_t,\, z_t \mid f^t(z_1, \cdots, z_{t-1})).\end{aligned} \tag{4}$$

We set

$$W^n(y^n \mid f) = \sum_{z^n \in \mathcal{Z}^n} W^n(y^n,\, z^n \mid f) \tag{5}$$

and describe now the feedback codes with randomized encoding strategies, that is, elements of $\mathcal{P}(\mathcal{F}_n)$, the set of probability distributions on $\mathcal{F}_n$.

Manuscript received February 18, 1994; revised November 27, 1994. The work of Z. Zhang was supported in part by the National Science Foundation under Grant NCR-9205265 and by the University of Bielefeld under SFB 343, Diskrete Strukturen in der Mathematik. The material in this paper was presented at the International Symposium on Information Theory and Its Applications (ISITA), Sidney, Australia, November 20–24, 1994; and in part at the IEEE International Symposium on Information Theory, Trondheim, Norway, June 27–July 1, 1994.

R. Ahlswede is with the Universität Bielefeld, Fakultät für Mathematik, Postfach 100131, 33501 Bielefeld, Germany.

Z. Zhang is with the Communication Sciences Institute, University of Southern California, Los Angeles, CA 90089-2565 USA.

IEEE Log Number 9412180.

Definition 1: A randomized (n, N, λ) IDF code for W is a system

$$\{(Q_F(\cdot \mid i), \mathcal{D}_i) : i = 1, \cdots, N\}$$

with

$$Q_F(\cdot \mid i) \in \mathcal{P}(\mathcal{F}_n),\ \mathcal{D}_i \subset \mathcal{Y}^n, \qquad \text{for } i \in \{1, \cdots, N\}$$

and

$$\sum_{g \in \mathcal{F}_n} Q_F(g \mid i) W^n(\mathcal{D}_i \mid g) \geq 1 - \lambda$$
$$\sum_{g \in \mathcal{F}_n} Q_F(g \mid i) W^n(\mathcal{D}_j \mid g) \leq \lambda \tag{6}$$

for all $i, j \in \{1, \cdots, N\}$ and $i \neq j$.

Let $N_F(n, \lambda)$ be the maximum integer N for which a randomized (n, N, λ) IDF code for W exists. We also use

$$\tilde{W}(y \mid x) = \sum_z W(y, z \mid x)$$

and call $\{\mathcal{X}, \tilde{W}, \mathcal{Y}\}$ the main channel. Our goal is to determine the double exponential growth of $N_F(n, \lambda)$. Insofar we have the following result.

Theorem 1: If the transmission capacity C of the main channel $\tilde{W}$ is positive, then we have for all $\lambda \in \left(0, \frac{1}{2}\right)$

$$\liminf_{n \to \infty} \frac{1}{n} \log \log N_F(n, \lambda) \geq \max I(XU \wedge Y) \tag{7}$$

where the maximum is taken over all joint distributions P_{XYZU} with

$$P_{XYZU}(x, y, z, u) = p(x) W(y, z \mid x) q(u \mid x, z)$$

satisfying

$$I(U \wedge Z \mid XY) < I(X \wedge Y).$$

Furthermore

$$\limsup_{n \to \infty} \frac{1}{n} \log \log N_F(n, \lambda) \leq \max I(XZ \wedge Y), \tag{8}$$

where the maximum is taken over all joint distributions P_{XYZ} with

$$P_{XYZ}(x, y, z) = p(x) W(y, z \mid x).$$

Remarks:

1) This theorem implies the results of [3] and [4]. To see this, observe that in the case without feedback $Z = 0$ and both bounds equal Shannon's transmission capacity. This is the result of [3]. In the complete feedback case we have $Z = Y$ and both bounds equal the maximum entropy $H(Y)$. This is the result of [4]. Therefore, Theorem 1 can be viewed as a unification of the results of [3] and [4].

2) A challenging task is to close the gap between the two bounds. We guess that the lower bound is tight, however, a converse proof technique more powerful than those of [3] and [5] is needed!

The second problem concerns identification via a wiretap channel. This channel was introduced by Wyner [1]. It can be viewed as a probabilistic model for cryptography.

The channel has two outputs. One is for the legitimate receiver and the other, which is a degraded version of the first output, is for the wiretapper, The goal of the communication is to send messages to the legitimate receiver while the wiretapper must be kept ignorant. A more general version of the wiretap channel was studied in [2], where the assumption that the output for the wiretapper, is a degraded version of the output for the legitimate receiver is dropped. We address here right away this general model. As defined in [2], a wiretap channel is a quintuple

$$\{\mathcal{X}, W, V, \mathcal{Y}, \mathcal{Z}\} \tag{9}$$

where $\mathcal{X}$ is the input alphabet, $\mathcal{Y}$ is the output alphabet for the legitimate receiver, $\mathcal{Z}$ is the output for the wiretapper, $W = \{W(y \mid x) : x \in \mathcal{X},\ y \in \mathcal{Y}\}$ is the channel transmission matrix, whose output is available to the legitimate receiver, and $V = \{V(z \mid x) : x \in \mathcal{X},\ z \in \mathcal{Z}\}$ is the channel transmission matrix, whose output is available to the wiretapper. The channel is assumed to be memoryless, that is, the conditional probabilities of the output word y^n and z^n given the input word x^n are

$$W^n(y^n \mid x^n) = \prod_{t=1}^{n} W(y_t \mid x_t)$$

and

$$V^n(z^n \mid x^n) = \prod_{t=1}^{n} V(z_t \mid x_t).$$

In the classical transmission problem, an (n, M, ϵ)-code for the wiretap channel is defined as a system

$$\{(c_i, \mathcal{D}_i) : 1 \leq i \leq M\} \tag{10}$$

where for all i, $c_i \in \mathcal{P}(\mathcal{X}^n)$ are the codewords and $\mathcal{D}_i \subset \mathcal{Y}^n$ are the disjoint decoding sets. It is required that for any i

$$\lambda_i \triangleq W^n(\mathcal{D}_i^c \mid c_i) \leq \epsilon \tag{11}$$

and if X^n has uniform distribution over $\{c_i : 1 \leq i \leq M\}$, then

$$\frac{1}{n} I(X^n \wedge Z^n) \leq \epsilon. \tag{12}$$

The secrecy capacity of the wiretap channel is defined as the maximum rate of any code which satisfies these conditions. Formally, let

$$M(n, \epsilon) = \max\{M : \exists \text{ an } (n, M, \epsilon) \text{ code}\} \tag{13}$$

then the secrecy capacity of the wiretap channel is defined as

$$C_s = \max\{R : \forall \epsilon > 0,\ \exists n(\epsilon) \text{ such that for } n \geq n(\epsilon)\ M(n, \epsilon) \geq 2^{n\mathcal{R}}\}. \tag{14}$$

The secrecy capacity of the general wiretap channel was determined in [2]. It is

$$C_s = \max_{U \to X \to YZ} I(U \wedge Y) - I(U \wedge Z). \tag{15}$$

The problem of identification via this channel in the sense of [3] can be formulated as follows: For any finite set A let $\mathcal{P}(A)$ stand for the set of all probability distributions on A.

Definition 2: A randomized (n, N, λ)-identification code for the wiretap channel is a system

$$\{(Q(\cdot \mid i), \mathcal{D}_i): 1 \le i \le N\} \tag{16}$$

where, for all i, $Q(\cdot \mid i) \in \mathcal{P}(\mathcal{X}^n)$ and $\mathcal{D}_i \subset \mathcal{Y}^n$, which satisfies the following three conditions:

1) for all i

$$\sum_{x^n \in \mathcal{X}^n} Q(x^n \mid i) W^n(\mathcal{D}_i \mid x^n) \ge 1 - \lambda \tag{17}$$

2) for all pair (i, j) with $i \ne j$

$$\sum_{x^n \in \mathcal{X}^n} Q(x^n \mid j) W^n(\mathcal{D}_i \mid x^n) \le \lambda \tag{18}$$

and

3) for any pair (i, j) with $i \ne j$ and any $\mathcal{V} \subset \mathcal{Z}^n$

$$\sum_{x^n \in \mathcal{X}^n} Q(x^n \mid j) V^n(\mathcal{V} \mid x^n) + \sum_{x^n \in \mathcal{X}^n} Q(x^n \mid i) V^n(\mathcal{V}^c \mid x^n) \ge 1 - \lambda. \tag{19}$$

In contrast to the transmission problem, the decoding sets for the identification problem are not necessarily disjoint.

Condition 3 enforces that the wiretapper is kept with his error probability close to $\frac{1}{2}$. This is the highest possible value, because the wiretapper could just accept an i of his interest with probability $\frac{1}{2}$. Mathematically, Condition 3 means, of course, that the output distributions for the wiretap channel are almost the same for any two input distributions $Q(\cdot \mid i)$ and $Q(\cdot \mid j)$.

The maximum N for which a randomized (n, N, λ)-identification code exists is denoted by $N(n, \lambda)$. Define the secrecy identification capacity of the wiretap channel by letting

$$C_{si} = \max\{R: \forall \lambda > 0,\ \exists n(\lambda) \text{ such that for } n \ge n(\lambda)\ \ N(n, \lambda) \ge 2^{2^{nR}}\}.$$

The main result on this problem is the following:

Theorem 2 (Dichotomy Theorem): Let C be the Shannon capacity of the channel W and let C_s be the secrecy transmission capacity of the wiretap channel, then

$$C_{si} = C, \qquad \text{if } C_s > 0 \tag{20}$$

and

$$C_{si} = 0, \qquad \text{if } C_s = 0. \tag{21}$$

Remark 3: This result has solved the second problem. Still it may be of interest to know whether the strong converse holds.

The paper is organized as follows: In Section II we discuss the strong converse proof of Theorem 1 for the channel without feedback from [5], which will be used repeatedly throughout the paper, and we prove Theorem 1. Section III is devoted to the proof of Theorem 2.

II. Proof of Theorem 1

In this section we prove Theorem 1. The converse part is proved in the first subsection, the direct part is proved in the second subsection, and deterministic IDF codes are briefly discussed in the last subsection.

A. Converse Part of Theorem 1

For identification via channels without feedback a so-called soft converse was proved in the original paper [3]. The method was refined and strengthened in [5] to give a proof of the strong converse theorem. The techniques used in [3] and [5] are needed in this section as well as for wiretap channels. Although the proofs in [5] were already simplified compared with those in [3], they are still too long and too complicated to reproduce all the details here. Therefore, in this section, we briefly review some of the key steps and key definitions from [5] and present modifications necessary for our purposes. The details can be found in [5].

We start with a review of some definitions. A probability distribution Q on $\mathcal{A}$ is called an n-type if for any $a \in \mathcal{A}$, $Q(a) \in \{\frac{1}{n}, \cdots, \frac{i}{n}, \cdots, 1\}$. Let Γ be the set of all possible n-types. For any $Q \in \Gamma$ let

$$T_Q = \left\{a^n: \forall a \in \mathcal{A},\ \frac{|\{i: a_i = a\}|}{n} = Q(a)\right\}. \tag{22}$$

Let B be a subset of $\mathcal{A}^n$ and let Q_B denote the uniform distribution on B. Finally, an ID code $\{(Q_i, \mathcal{D}_i): 1 \le i \le M\}$ is called homogeneous if for every $P \in \Gamma$

$$Q_1(T_P) = \cdots = Q_M(T_P).$$

We state the first auxiliary result from [5].

Proposition 1: For every (n, N, λ)-ID code, $\delta > 0$, $\lambda' > \lambda$ and all sufficiently large n, there exists a homogeneous (n, N', λ')-ID code satisfying $N' > N \exp\{-\delta n (n+1)^{|\mathcal{X}|}\}$, where $\mathcal{X}$ is the input alphabet.

For a distribution Q on $\mathcal{X}^n$ and every $P \in \Gamma$, define

$$Q^P(x^n) = \frac{Q(x^n)}{Q(T_P)}.$$

An ID code is called M-regular if for every $P \in \Gamma$ and all i, Q_i^P is of M-type.

Proposition 2: A homogeneous M-regular (n, N, λ)-ID code with $\lambda < \frac{1}{2}$ satisfies

$$\log N \le n(n+1) M \log |\mathcal{X}|.$$

The main result of [5], that is, the strong converse for the channels without feedback, follows easily from the following result.

Proposition 3: For every homogeneous (n, N, λ)-ID code, $\lambda' > \lambda$, $\gamma > 0$, and for all sufficiently large n, there exists a homogeneous $\exp\{nC + n\gamma\}$-regular (n, N, λ')-ID code, where C is the Shannon channel capacity.

In the proof of this proposition the following lemma of [5] is needed.

Lemma 1: Let $P \in \Gamma$ and let Q be a probability distribution on T_P. For every $\epsilon \in [0, \epsilon_0]$, $\delta \in [0, \delta_0]$ and for all sufficiently large n, there exists an $\exp\{nC + n\gamma\}$-type distribution Q defined on T_P, where C is the Shannon channel capacity, such that for every $\mathcal{D} \subset \mathcal{B}^n$, where $\mathcal{B}$ is the output alphabet,

$$\tilde{Q}W^n(\mathcal{D}) \le (1+\epsilon)(1-e^{-n\delta})^{-1}QW^n(\mathcal{D}) + e^{-n\delta} \tag{23}$$

$$\tilde{Q}W^n(\mathcal{D}) \ge (1-\epsilon)(1-e^{-n\delta})QW^n(\mathcal{D}) - e^{-n\delta} \tag{24}$$

where W is the channel transition probability matrix and where $\gamma = \rho(\delta)$, and $\rho\colon [0, \delta_0] \to R^+$ is a continuous strictly increasing function such that $\rho(0) = 0$.

By checking the proof of this lemma in [5] we can find that actually the following stronger result was proved:

Lemma 2: Let $P \in \Gamma$ and let Q be a probability distribution on T_P. For every $\epsilon \in [0, \epsilon_0]$, $\delta \in [0, \delta_0]$, let $\overline{U} = \{U_1, \cdots, U_{M'}\}$ be a random code having independent codewords and with codeword distribution Q. For every $R > I(P, W)$, $M' = e^{nR+n\gamma}$, where γ is defined in Lemma 1, the probability of the event that the following conditions are satisfied approaches 1 as n goes to infinity: For all $\mathcal{D} \subset \mathcal{B}^n$

$$\tilde{Q}W^n(\mathcal{D}) \le (1+\epsilon)(1-2^{-n\delta})^{-1}QW^n(\mathcal{D}) + e^{-n\delta} \tag{25}$$

$$\tilde{Q}W^n(\mathcal{D}) \ge (1-\epsilon)(1-e^{-n\delta})QW^n(\mathcal{D}) - e^{-n\delta} \tag{26}$$

where $\tilde{Q}$ is the uniform distribution on $\overline{U}$.

Since the original proof of Lemma 1 is extremely complicated, instead of copying it step by step we just point out the modifications needed to reach the current conclusions. In this new version, the lemma is strengthened in two points:

1) C is replaced by any $R \ge I(P, W)$

2) the existence of such a code is replaced by the conclusion that the random code satisfies (25) and (26) with probability approaching 1.

The first conclusion can be justified by noticing that [5, eq. (32)] is really unnecessary for the lemma. We need only

$$\sup_{V \in \Gamma_\delta^P} I(P, V) + \delta \le I(P, W) + \rho(\delta) \le R + \rho(\delta).$$

This is so, because we are considering a fixed P anyway. The second conclusion comes from the following refinement of [5, Lemma 5].

Lemma 3: Let $(\tilde{u}_1, \cdots, \tilde{u}_M)$ be the realization of the i.i.d. random variables $(U_1, \cdots, U_M)$ with common distribution Q. Let $\mathcal{E}$ be the event that the following conditions are satisfied:

$$\frac{1}{M}\sum_{i=1}^{M} 1\{\tilde{u}_i \in H_V^P(y^n)\} \le (1+\epsilon)Q(H_V^P(y^n)), \qquad \text{for every } y^n \in G_V^P$$

$$\frac{1}{M}\sum_{i=1}^{M} 1\{\tilde{u}_i \in H_V^P(y^n)\} \ge (1-\epsilon)Q(H_V^P(y^n)), \qquad \text{for every } y^n \in G_V^P$$

and

$$\frac{1}{M}\sum_{i=1}^{M} W_V^P((G_V^P)^c \mid \tilde{u}_i) \le e^{-n\delta/3}.$$

Then we have

$$\Pr(\mathcal{E}) \le e^{-n\delta/3} + 2e^{-(\delta^2/3)e^{n\delta}}.$$

A careful check of the proof of [5, Lemma 5] shows that the conclusion is what is actually proved, although the statement of the lemma is slightly weaker.

In our proofs we use these definitions and results.

The converse part can be proved by following the argument in [5] with certain modifications. We present in this subsection only the modifications without going into all the details. Since the proofs of the two converses are very similar, pointing out these modifications is good enough for the readers to complete the proof by going through the proof in [5].

Let f be a feedback coding function. With the function f a set of pairs $\mathcal{S}_f = \{(x^n, z^n)\colon f(z^n) = x^n\}$ is associated. The probability of a pair $(x^n, z^n) \in \mathcal{S}_f$ is

$$W^n(z^n \mid x^n) = \sum_{y^n} W^n(y^n, z^n \mid x^n).$$

This gives a distribution on the set $\mathcal{X}^n \times \mathcal{Z}^n$. We denote it by $P_f(x^n, z^n)$. Let $Q(\cdot \mid i) \in \mathcal{P}(\mathcal{F}_n)$ be the distribution of the user i. This induces a distribution on $\mathcal{X}^n \times \mathcal{Z}^n$ defined as follows:

$$P_i(x^n, z^n) = \sum_{f \in \mathcal{F}_n} Q(f \mid i)P_f(x^n, z^n).$$

An n-type P on $\mathcal{X} \times \mathcal{Z}$ is called ϵ-typical if for any $x \in \mathcal{X}$ and any $z \in \mathcal{Z}$

$$\left| \frac{P(x, z)}{\sum_{z'} P(x, z')} - W(z \mid x) \right| \le \epsilon.$$

Let $\mathcal{P}_n^\epsilon$ be the set of all possible ϵ-typical n-types, then we have from the weak law of large numbers

$$\lim_{n \to \infty} P_i \left(\bigcup_{P \in \mathcal{P}_n^\epsilon} T_P \right) = 1.$$

The idea of the proof is the following: a feedback code $\{(Q(\cdot \mid i), \mathcal{D}_i)\colon 1 \le i \le N\}$ induces an identification code without feedback for the channel from (x, z) to y with transition probability

$$\tilde{W}(y \mid x, z) = \frac{W(y, z \mid x)}{\sum_{y'} W(y', z \mid x)}$$

of the form

$$\{(P_i, \mathcal{D}_i)\colon 1 \le i \le N\}.$$

Therefore, the proofs from [5] can be easily modified and applied to this induced code. The following proposition is modified from [5, Proposition 5].

Proposition 4: For every (n, N, λ)-feedback identification code and $\lambda' > \lambda$, $\gamma > 0$, $\delta > 0$ there exists a homogeneous $\exp\{nT + n\gamma\}$-regular (n, N', λ')-ID code, where $N' > N\exp\{-\delta n(n+1)^{|\mathcal{X}||\mathcal{Z}|}\}$, and $T = \max_p I(XZ \wedge Y)$ where the joint distribution P_{XYZ} satisfies $P_{XYZ}(x, y, z) = p(x)W(y, z \mid x)$.

This proposition is proved by an argument as that used in the proof of [5, Proposition 5]. The only difference is that [5, Lemma 1] is now replaced by Lemma 2 of this paper. Therefore C is replaced by

$$\max_{P \in \mathcal{P}_n^\epsilon} I(P, \tilde{W}) = T + \nu$$

where ν is a continuous function of ϵ satisfying $\nu(0) = 0$.

The converse is proved now by the same argument as in [5] with [5, Proposition 5] replaced by this new proposition.

B. Proof of the Direct Part of Theorem 1

The proof of the direct part of Theorem 1 is based on two ideas.

The first one is the idea presented in [4, Section III], where the identification code is constructed by means of two fundamental codes. One code is of block length n and the other one is of block length m, which is much smaller than n. The task of the first code is to set up a common random experiment. The result of the experiment, which is known with high probability to both, the encoder and the decoder, serves as a "public key." According to this key a codeword of the second code is transmitted in the second step. Two different users use the same codeword for the same public key with very small probability. Therefore, the goal of identification is achieved.

The second idea is the well-known superposition coding scheme introduced in [9]. In this coding scheme, there are K steps. In each step, a codeword is sent to transmit a new message as well as to resolve an uncertainty left over from the previous step.

For given $\delta > 0$ and $\epsilon > 0$, let P_{XYZU} be a probability distribution $P_{XYZU}(x, y, z, u) = p(x)W(y, z \mid x)q(u \mid x, z)$ that achieves $\max I(U \wedge Z \mid X)$ under the constraint

$$I(U \wedge Z \mid XY) \leq I(X \wedge Y) - \delta.$$

We construct three codes of block length n using p and q of this form.

Code $\mathcal{C}_1$: Code $\mathcal{C}_1$ is an $(n, M_n, 2^{-n\alpha})$ channel code for the channel with

$$W(y \mid x) = \sum_z W(y, z \mid x).$$

The codewords are assumed to be in T_p, where we assume without loss of generality that P is an n-type. The cardinality of the code is $M_n = 2^{nI(X \wedge Y) - \epsilon n}$, where ϵ is the given positive number which is assumed to be sufficiently small. Let $\{\mathcal{D}_i^{(n)} : 1 \leq i \leq M\}$ be the decoding regions of the codewords of $\mathcal{C}_1$ with maximum decoding error at most $2^{-\alpha n}$, where $\alpha > 0$ is a continuous function of ϵ satisfying $\alpha(0) = 0$.

Code Family $\mathcal{C}_2(c)$: Code family $\mathcal{C}_2(c) \subset \mathcal{U}^n$, where $\mathcal{U}$ is the alphabet of the random variable U, is a family of source codes indexed by the codewords $c \in \mathcal{C}_1$. This family of codes are required to satisfy the following conditions:

1) Given $c \in \mathcal{C}_1$ the codewords in $\mathcal{C}_2(c)$ are jointly ϵ-typical with c with respect to the joint distribution P_{XU}, a marginal of P_{XYZU}.

2) The cardinalities of the codes are $N_n = 2^{nI(U \wedge Z|X) + \epsilon n}$ for all $c \in \mathcal{C}_1$.

3) For each $c \in \mathcal{C}_1$, there exists a mapping

$$f_c \colon \mathcal{Z}^n \rightarrow \mathcal{C}_2(c)$$

satisfying

a) If $f_c(z^n) \neq 0$, then $f_c(z^n)$, c and z^n are jointly ϵ-typical.

b) $\Pr(f_c(z^n) = 0 \mid c) \leq 2^{-n\beta}$, where β is a continuous function of ϵ satisfying $\beta(0) = 0$.

Code Family $\mathcal{C}_3(c)$: The code family $\mathcal{C}_3(c)$ consists of an integer set $\{1, \cdots, L_n\}$, where $L_n = 2^{n(I(Z \wedge U|XY) + n\gamma)}$ and γ is a continuous function of ϵ satisfying $\gamma(0) = 0$, and two mappings defined as follows:

$$\Phi_c : \mathcal{C}_2(c) \rightarrow \{1, \cdots, L_n\}$$

and

$$\Psi_c : \mathcal{Y}^n \times \{1, \cdots, L_n\} \rightarrow \mathcal{C}_2(c)$$

satisfying

$$\Pr(\Psi_c(Y^n, \Phi_c(f_c(Z^n))) \neq f_c(Z^n) \mid c) \leq 2^{-n\sigma}$$

where σ is a continuous function of ϵ with $\sigma(0) = 0$.

Proof of the existence of the codes (code families): The existence of the code $\mathcal{C}_1$ is based on the channel coding theorem with maximum error criterion ([10]).

The existence of the code family $\mathcal{C}_2(c)$ is proved by the random coding method. Since the method is classical, we give only a brief outline of the proof. The code is selected randomly according to the distribution

$$r^n(u^n \mid c) = \sum_{y^n, z^n} q^n(u^n \mid c, z^n)W^n(y^n, z^n \mid c).$$

The N_n codewords are selected independently. The mapping f_c is defined by using joint ϵ-typicality as follows:

1) If there exists a unique codeword $c_2(c, i)$ in $\mathcal{C}_2(c)$ which is jointly ϵ-typical with z^n and c, then let $f_c(z^n) = i$,

2) otherwise, let $f_c(z^n) = 0$.

The properties of f_c are proved by using the properties of the joint ϵ-typical sequences. These proofs are standard and therefore omitted.

The existence of the code family $\mathcal{C}_3(c)$ is proved by using the source coding theorem with side information and by noticing the following fact. Since the joint distribution of c, Y^n, U^n, Z^n, and $f_c(Z^n)$ are given by the joint distribution of

X^n, Y^n, Z^n, U^n, the code $\mathcal{C}_2(c)$, and the mapping f_c, then

$$\begin{aligned} &H(f_c(Z^n) \mid c, Y^n) \\ &\quad = I(Z^n \wedge f_c(Z^n) \mid c, Y^n) \\ &\quad = H(Z^n \mid c, Y^n) - H(Z^n \mid c, Y^n, f_c(Z^n)) \\ &\quad = H(Z^n \mid c, Y^n) - H(Z^n \mid c, Y^n, U^n, f_c(Z^n)) \end{aligned}$$

(where U^n is the codeword of $\mathcal{C}_2(c)$ whose index is the value of $f_c(Z^n)$)

$$\begin{aligned} &\geq H(Z^n \mid c, Y^n) - H(Z^n \mid c, Y^n, U^n) \\ &\quad = \sum_i H(Z_i \mid x_i, Y_i) - H(Z_i \mid x_i, Y_i, U_i) \\ &\quad = nI(Z \wedge U \mid X, Y) + \beta n \end{aligned}$$

where β goes to zero as ϵ goes to zero. In the last step of the derivation we used the typicality of the codewords. Applying the source coding theorem with side information (if necessary, we may repeat the same code N times and use nN in place of n) to this case gives the existence of the code family $\mathcal{C}_3(c)$.

Using these three codes (code families), the coding scheme can be described. It includes two steps. In the first step, the sender and the receiver set up with high probability a common random experiment. In the second step, based on the result of the common random experiment, the sender sends a codeword to the receiver.

We formulate the two steps as follows:

1) The coding is done in K blocks. Each block is of length n. The code $\mathcal{C}_1$ is partitioned into $L_n = 2^{nI(U\wedge Z|XY)+\gamma n}$ subcodes of equal size (roughly) $B_n = 2^{n(I(X\wedge Y)-I(U\wedge Z|X,Y)-\gamma-\epsilon)}$, which are denoted by $\mathcal{C}_1^m$ for $m = 1, \cdots, L_n$. The codewords of the subcodes are indexed by the numbers in $\{1, \cdots, B_n\}$. Since $I(X \wedge Y) > I(U \wedge Z \mid X, Y) + \delta$, this is possible for ϵ small enough. We send a fixed codeword, say c_1 in $\mathcal{C}_1$, in the first block. In the second block, based on the feedback signal Z_1^n in the first block, we send a $c_2 \in \mathcal{C}_1^m$ where $m = \Phi_{c_1}(f_{c_1}(Z_1^n))$ is determined by the channel and where c_2 is selected from the code $\mathcal{C}_1^m$ randomly with respect to the uniform distribution. In the following steps i for $i = 3$ to K, if the feedback signal in the previous step is Z_{i-1}^n, then the sender sends $c_i \in \mathcal{C}_1^{m_i}$ where $m_i = \Phi_{c_{i-1}}(f_{c_{i-1}}(Z_{i-1}^n))$. c_i is selected randomly with respect to the uniform distribution in $\mathcal{C}_1^{m_i}$. The codewords $c_1, \cdots, c_K$ can be correctly decoded with probabilities at least $1 - K2^{-n\alpha}$. Then the codewords of the second code $\{f_{c_1}(Z_1^n), \cdots, f_{c_{K-1}}(Z_{K-1}^n)\}$ can be recovered with probability at least $1 - (K-1)2^{-n\sigma}$ under the condition that the codewords from the code $\mathcal{C}_1$ are correctly decoded. The overall misdecoding probability is at most

$$P_e = K(2^{-n\alpha} + 2^{-n\gamma}).$$

This means that with probability at least $1 - P_e$ the sender and the receiver have a common knowledge of $\{f_{c_1}(Z_1^n), \cdots, f_{c_{K-1}}(Z_{K-1}^n)\}$ and the indices b_i of the codewords c_i in their corresponding subcodes $\mathcal{C}_1^m$ of the code $\mathcal{C}_1$, which are numbers from the set $\{1, \cdots, B_n\}$, at the end of the first K blocks. They are viewed as the result of the common random experiment.

2) Let

$$\mathcal{F} = \{F : F : \{1, \cdots, N_n\}^{K-1} \times \{1, \cdots, B_n\}^K \to \mathcal{C}_1\}. \quad (27)$$

Each user is assigned a mapping in $\mathcal{F}$. Let F_j be the mapping assigned to the user j. Once $\{f_{c_1}(Z_1^n), \cdots, f_{c_{K-1}}(Z_{K-1}^n)\}$ and $\{b_1, \cdots, b_K\}$ from the first K steps are available to the sender (and with probability at least $1 - P_e$ correctly to the receiver), the user j selects a codeword

$$F_j(f_{c_1}(Z_1^n), \cdots, f_{c_{K-1}}(Z_{K-1}^n), b_1, \cdots, b_K) \in \mathcal{C}_1$$

and sends it through the channel. This codeword can be decoded correctly with probability at least $2^{-n\alpha}$.

We now prove that, if the user number satisfies a certain condition, then there exists mappings F_j for the users such that the two kinds of error probabilities of the code described above satisfy the requirement of the identification code.

Obviously, the misrejection probability is at most $1 - (K + 1)(2^{-n\alpha} + 2^{-n\gamma})$, which goes to zero as n goes to infinity for a fixed K.

The misacceptance probability can be estimated as follows: we assume that the mapping is selected according to the uniform distribution on $\mathcal{F}$ and the selection for different users are independent. Let F_0 be the mapping assigned to the user to be identified, let F_i be the mapping assigned to the user i. For a particular $\overline{v} = (f_{c_1}(Z_1^n), \cdots, f_{c_{K-1}}(Z_{K-1}^n), b_1, \cdots, b_K)$

$$F_0(\overline{v}) = F_i(\overline{v})$$

with probability $2^{-nI(X\wedge Y)+n\epsilon} = M_n^{-1}$. The misacceptance probability, when the user is i, is greater than $\lambda + p_e$ (where p_e is the probability that the receiver and the sender cannot reach a common result of the random experiment and which goes to zero as n goes to infinity) with probability at most

$$N_n^{K-1} B_n^K (1 - 2^{-nI(X\wedge Y)+n\epsilon})^{N_n^{K-1} B_n^K - G} \cdot (2^{-nI(X\wedge Y)+n\epsilon})^G \binom{N_n^{K-1} B_n^K}{G}$$

where

$$G = \max\{|\mathcal{V}| : \Pr(\overline{v} \in \mathcal{V}) \leq \lambda\}.$$

Since any set $\mathcal{V}$ with cardinality at most

$$2^{-2Kn\epsilon} N_n^{K-1} B_n^K = 2^{n(K-1)I(U\wedge Z|X)-n\epsilon} B_n^K$$

has a vanishing probability as n goes to infinity, we have

$$G \geq 2^{-2Kn\epsilon} N_n^{K-1} B_n^K.$$

Therefore, when $2K\epsilon < I(X \wedge Y) - \epsilon$ this probability has an upper bound

$$\begin{aligned} &2^{-2^{n(K-1)I(U\wedge Z|X)+nK(I(X\wedge Y)-I(U\wedge Z|X,Y))-n(K+1)\epsilon-nK\gamma+o(nK)}} \\ &\qquad \leq 2^{-2^{n(K-1)I(XU\wedge Y)-n(K+1)\epsilon-nK\gamma+o(nK)}}. \end{aligned}$$

When user i is to be identified, the probability that there exists a user $j \neq i$ having misacceptance probability at least $\lambda + p_e$ is at most

$$1 - \left(1 - 2^{-2^{n(K-1)I(XU\wedge Y)-n(K+1)\epsilon-nK\gamma+o(nK)}}\right)^M.$$

This goes to zero when

$$M \le 2^{2^{n(K-1)I(XU\wedge Y)-n(K+1)\epsilon-nK\gamma+o(nK)}-\epsilon n}.$$

If we delete users for whom there exists at least one different user having misacceptance probability at least $\lambda + p_e$, then the number of users deleted is in average a vanishing portion of all M users. Therefore, there exists a set of mappings for M' users out of the M users, where

$$M' = 2^{2^{n(K-1)I(XU\wedge Y)-n(K+1)\epsilon-nK\gamma+o(nK)}-2\epsilon n}$$

such that the subcode of these users satisfies all requirements of the identification code. The rate of the code is at least

$$\frac{K-1}{K+1} I(XY \wedge Y) - 2\epsilon - \gamma.$$

For large K, this is greater than $I(XU \wedge Y) - 3\epsilon - \gamma$. Letting ϵ go to zero the direct part of Theorem 1 is proved.

C. Deterministic Identification Codes

Another result of [4] is for deterministic feedback identification codes. Actually, the same concept can be defined for channels with noisy feedback. In this subsection, we present only the definitions and results for this concept without detailed proofs. These results are proved by the method used for the randomized identification code with some modifications.

Definition 3: A deterministic (n, N, λ) IDF code for W is a system

$$\{(f_i, \mathcal{D}_i) : i = 1, \cdots, N\}$$

with

$$f_i \in \mathcal{F}_i,\ \mathcal{D}_i \subset \mathcal{Y}^n, \qquad \text{for } i \in \{1, \cdots, N\}$$

and

$$W^n(\mathcal{D}_i \mid f_i) \ge 1 - \lambda \qquad W^n(\mathcal{D}_j \mid f_i) \le \lambda \tag{28}$$

for all $i, j \in \{1, \cdots, N\}$ and $i \ne j$.

Let $N_f(n, \lambda)$ be the maximum integer N for which a deterministic (n, N, λ) IDF code exists.

Here are our results for this quantity.

Theorem 3: If the transmission capacity C of W is positive, then we have for all $\lambda \in \left(0, \frac{1}{2}\right)$

$$\liminf_{n\to\infty} \frac{1}{n} \log\log N_f(n, \lambda) \ge \max I(Z \wedge U \mid X) \tag{29}$$

where the maximum is over all joint distributions P_{XYZU} with

$$P_{XYZU}(x, y, z, u) = p(x)W(y, z \mid x)q(u \mid x, z)$$

satisfying

$$I(U \wedge Z \mid XY) < I(X \wedge Y).$$

Furthermore

$$\limsup_{n\to\infty} \frac{1}{n} \log\log N_f(n, \lambda) \le \min\{\max I(XZ \wedge Y), \max H(Z \mid X)\} \tag{30}$$

where the maximum is taken over all joint distributions P_{XYZ} with

$$P_{XYZ}(x, y, z) = p(x)W(y, z \mid x).$$

III. Proof of Theorem 2

In this section we prove Theorem 2. The direct part is proved in the first three subsections and the converse part is proved in the last subsection.

A. Preparations for the Proof of the Direct Part

In the proof of the direct part of Theorem 2, we use a coding technique introduced in [4, Section III], where the identification code is constructed by means of two fundamental codes. This coding technique has been already used for channels with noisy feedback. By Shannon's coding theorem, we know that for every $\epsilon > 0$, $\epsilon < C$, where C is the Shannon capacity of the main channel W, there is a $\delta = \delta(\epsilon) > 0$ and an $n_0(\epsilon)$ such that for $n > n_0(\epsilon)$, there exists an $(n, M, 2^{-n\delta})$ code

$$\mathcal{C}_1 = \{(\tilde{c}_j, \tilde{\mathcal{C}}_j) : j = 1, \cdots, M\} \tag{31}$$

where $M = 2^{n(C-\epsilon)}$. This code serves as the first fundamental code which will be used in the construction of the identification code.

For the wiretap channel, in place of the second fundamental code, we use a code system which consists of a code of length m and a collection of subcodes of this code. This code system should satisfy certain conditions described later. To construct this code system, we use a random variable U jointly distributed with random variables X, Y, and Z. The joint distribution of these random variables is of the form

$$P_{UXYZ}(u, x, y, z) = q(u)r(x \mid u)W(y, z \mid x) \tag{32}$$

which satisfies the condition

$$I(U \wedge Y) > I(U \wedge Z). \tag{33}$$

The following proposition gives the existence of a code system which will be used in the construction of the identification code. Let

$$\tilde{W}(y \mid u) = \sum_x r(x \mid u)W(y \mid x).$$

$\tilde{W}$ is called the u, y-channel. Let

$$\tilde{V}(z \mid u) = \sum_x r(x \mid u)V(z \mid x).$$

$\tilde{V}$ is called the u, z-channel.

Proposition 5: For any $\epsilon > 0$ there exists a $\delta(\epsilon) > 0$ and an m_0 such that for any $m > m_0$ there exist an $(m, M', 2^{-m\delta})$ code

$$\mathcal{C}_2 = \{(c'_i, \mathcal{D}'_i) : 1 \le i \le M'\} \tag{34}$$

for the u, y-channel $\tilde{W}$, where $M' = 2^{m(I(U\wedge Y)-\epsilon)}$, and $L = 2^{m\epsilon}$ subcodes

$$\mathcal{L} = \{\mathcal{C}_i^* : i = 1, \cdots, L\} \tag{35}$$

of the code $\mathcal{C}_2$ with a common cardinality $M^* = 2^{m(I(U\wedge Z)+\epsilon)}$ having the following two properties:

1) The number of common codewords of any two different subcodes is at most ϵM^*.

2) Let Q_i be the uniform distribution on $\mathcal{C}_i^*$. Then for every pair i and $j: i \neq j$

$$D(Q_i\tilde{V}^m \| Q_j\tilde{V}^m) \leq \epsilon.$$

This proposition will be proved in Subsection III-C.

B. Proof of the Direct Part of Theorem 2

Using these two fundamental codes, we can construct the identification code as follows:

Let $\{1, \cdots, L\}$ be the index set of $\mathcal{L}$, the set of subcodes of the second fundamental code $\mathcal{C}_2$. We consider mappings of the form

$$\phi: \mathcal{C}_1 \rightarrow \{1, \cdots, L\}. \tag{36}$$

Let Φ be the set of all possible mappings ϕ. The Hamming distance of two mappings ϕ and ψ is defined as the number of codewords of $\mathcal{C}_1$ at which ϕ and ψ have different values. It can easily be seen that we can construct by the greedy algorithm a set of mappings of cardinality at least

$$\begin{aligned} N &= \frac{L^M}{\sum\limits_{k=\epsilon M}^{M} \binom{M}{k} (L-1)^{M-k}} \\ &\geq \frac{2^{MD(\epsilon \| L^{-1})}}{M} \geq 2^{2^{nC-2\epsilon n}} \end{aligned}$$

where

$$D(\epsilon \| L^{-1}) = \epsilon \log \frac{\epsilon}{L^{-1}} + (1-\epsilon) \log \frac{1-\epsilon}{1-L^{-1}}$$

satisfying the property that the Hamming distance between any pair of different mappings in the set is at least $M - \epsilon M$. Let this set of mappings be

$$\Phi^* = \{\phi_i: 1 \leq i \leq N\} \tag{37}$$

which will be used in the construction of the identification codes. Let P be the uniform distribution over the code $\mathcal{C}_1$, q_i^* be the uniform distribution on the subcode $\mathcal{C}_i^*$, and let $Q_i^* = q_i^* r^m$, which is a distribution on the alphabet $\mathcal{X}^m$. Let mapping ϕ_i be assigned to user i, then the distribution $Q(\cdot \mid i)$ in the identification code is defined as follows: for $x^n \in \mathcal{X}^n$ and $x^m \in \mathcal{X}^m$

$$Q((x^n, x^m) \mid i) = P(x^n) Q^*_{\phi_i(x^n)}(x^m). \tag{38}$$

The decoding set $\mathcal{D}_i$ is defined as

$$\mathcal{D}_i = \bigcup_{t=1}^{M} \tilde{\mathcal{D}}_t \times \mathcal{D}^{\phi_i c(\tilde{c}_t)}$$

where

$$\mathcal{D}^{(s)} = \bigcup_{c' \in \mathcal{C}_s^*} \mathcal{D}'_{c'}$$

and where $\tilde{\mathcal{D}}_t$ was defined in (31) and $\mathcal{D}'_i$ when $c' = c'_i$ (which was defined in (34)). We now estimate the first and the second kinds of error probabilities of this code.

We have for user i, the first kind of error probability is

$$\begin{aligned} &\sum_{(x^n, x^m) \in \mathcal{X}^{n+m}} Q((x^n, x^m) \mid i) W^{n+m}(\mathcal{D}_i \mid x^n, x^m) \\ &\geq 1 - \Bigg(1 - \sum_{t=1}^{M} P(\tilde{c}_t) W^n(\tilde{\mathcal{D}}_t \mid \tilde{c}_t) + 1 \\ &\qquad - \sum_{t=1}^{M} P(\tilde{c}_t) \sum_{c' \in \mathcal{C}_{\Phi_i(\tilde{c}_t)}} q^*_{\phi_i(\tilde{c}_t)}(c') \tilde{W}^m(\mathcal{D}^{(\phi_i(\tilde{c}_i))} \mid c')\Bigg) \\ &\geq 1 - 2^{-n\delta} - 2^{-m\delta}. \end{aligned} \tag{39}$$

If $i \neq j$, then the second kind error probability for the user j, when user i is to be identified, is

$$\begin{aligned} &\sum_{(x^n, x^m) \in \mathcal{X}^{n+m}} Q((x^n, x^m) \mid j) W^n(\mathcal{D}_i \mid x^n, x^m) \\ &\leq 1 - \sum_{t=1}^{M} P(\tilde{c}_t) W^n(\tilde{\mathcal{D}}_t \mid \tilde{c}_t) + \sum_{t=1}^{M} P(\tilde{c}_t) \\ &\quad \cdot \sum_{c' \in \mathcal{C}_{\phi_j(\tilde{c}_t)}} q^*_{\phi_j(\tilde{c}_t)}(c') \tilde{W}^m(\mathcal{D}^{(\phi_i(\tilde{c}_t))} \mid c') \\ &\leq 2^{-n\delta} + \sum_{t: \phi_i(\tilde{c}_t) = \phi_j(\tilde{c}_t)} P(\tilde{c}_t) + \sum_{t: \phi_i(\tilde{c}_t) \neq \phi_j(\tilde{c}_t)} P(\tilde{c}_t) \\ &\quad \cdot \sum_{c' \in \mathcal{C}_{\phi_j(\tilde{c}_t)}} q^*_{\phi_j(\tilde{c}_t)}(c') \tilde{W}^m(\mathcal{D}^{(\phi_i(\tilde{c}_t))} \mid c') \\ &\leq 2^{-n\delta} + \epsilon + \sum_{t: \phi_i(\tilde{c}_t) \neq \phi_j(\tilde{c}_t)} P(\tilde{c}_t) \\ &\quad \cdot \sum_{c' \in \mathcal{C}_{\phi_j(\tilde{c}_t)}} q^*_{\phi_j(\tilde{c}_t)}(c') \sum_{c'' \in \mathcal{C}_{\phi_i(\tilde{c}_t)}} \tilde{W}^m(\mathcal{D}'_{c''} \mid c') \\ &\leq 2^{-n\delta} + \epsilon + \sum_{t: \phi_i(\tilde{c}_t) \neq \phi_j(\tilde{c}_t)} P(\tilde{c}_t) \\ &\quad \cdot \sum_{c' \in \mathcal{C}_{\phi_j(\tilde{c}_t)} \cap \mathcal{C}_{\phi_i(\tilde{c}_t)}} q^*_{\phi_j(\tilde{c}_t)}(c') \tilde{W}^m(\mathcal{D}'_{c'} \mid c') \\ &\quad + \sum_{t: \phi_i(\tilde{c}_t) \neq \phi_j(\tilde{c}_t)} P(\tilde{c}_t) \\ &\quad \cdot \sum_{c' \in \mathcal{C}_{\phi_j(\tilde{c}_t)}} q^*_{\phi_j(\tilde{c}_t)}(c') \tilde{W}^m((\mathcal{D}'_{c'})^c \mid c') \\ &\leq 2^{-n\delta} + 2\epsilon + 2^{-m\delta}. \end{aligned} \tag{40}$$

For a fixed $\lambda < \frac{1}{2}$, let n be sufficiently large and then m be sufficiently large and ϵ sufficiently small, the requirement for these two error probabilities in the definition of the identification code can be satisfied.

The next problem is to prove that the wiretapper cannot identify, that is, we need to prove (19).

We see that for any pair $i \neq j$

$$\begin{aligned} &D(Q(\cdot \mid i)V^{n+m} \| Q(\cdot \mid j)V^{n+m}) \\ &\quad = \sum_{\tilde{c} \in \mathcal{C}_1} P(\tilde{c}) D(Q_{\phi_i(\tilde{c})}V^m \| Q_{\phi_j(\tilde{c})}V^m) \\ &\quad \leq \sum_{\tilde{c} \in \mathcal{C}_1} P(\tilde{c})\epsilon = \epsilon. \end{aligned} \tag{41}$$

Therefore, for any region $\mathcal{V} \subset \mathcal{Z}^{n+m}$, denoting by V_i the distribution $Q(\cdot \mid i)V^{n+m}$ and by V_j the distribution $Q(\cdot \mid j)V^{n+m}$, we obtain

$$V_i(\mathcal{V}) \log \frac{V_i(\mathcal{V})}{V_j(\mathcal{V})} + V_i(\mathcal{V}^c) \log \frac{V_i(\mathcal{V}^c)}{V_j(\mathcal{V}^c)} \le D(V_i \| V_j) \le \epsilon.$$

Since

$$\begin{aligned} D(\alpha \| \beta) &= \alpha \log \frac{\alpha}{\beta} + (1-\alpha) \log \frac{1-\alpha}{1-\beta} \\ &= -\alpha \log \left(1 + \frac{\beta - \alpha}{\alpha}\right) \\ &\quad - (1-\alpha) \log \left(1 + \frac{\beta - \alpha}{1-\alpha}\right) \\ &\ge -\alpha \frac{\beta - \alpha}{\alpha} - (1-\alpha) \frac{\beta - \alpha}{1-\alpha} \\ &= 2(\beta - \alpha). \end{aligned}$$

Similarly

$$D(\beta \| \alpha) \ge 2(\alpha - \beta).$$

From

$$D(V_i \| V_j) \le \epsilon$$

and

$$D(V_j \| V_i) \le \epsilon$$

we obtain

$$V_i(\mathcal{V}) \log \frac{V_i(\mathcal{V})}{V_j(\mathcal{V})} + V_i(\mathcal{V}^c) \log \frac{V_i(\mathcal{V}^c)}{V_j(\mathcal{V}^c)} \le \epsilon$$

and

$$V_j(\mathcal{V}) \log \frac{V_j(\mathcal{V})}{V_i(\mathcal{V})} + V_j(\mathcal{V}^c) \log \frac{V_j(\mathcal{V}^c)}{V_i(\mathcal{V}^c)} \le \epsilon.$$

This implies

$$|V_i(\mathcal{V}) - V_j(\mathcal{V})| \le \frac{\epsilon}{2}.$$

We can see that the last inequality implies the last requirement for the identification code. The direct part of Theorem 2 is proved.

C. Proof of Proposition 5

To construct the second fundamental code with the required structure and properties, we use the random coding method. It is well known that there exists an $(m, M', 2^{-m\delta})$ code $\mathcal{C}_2$. Without loss of generality, we can assume that the distribution q of U is an m-type and $\mathcal{C}_2 \subset T_q$. To construct the random family of subcodes of this code $\{\mathcal{C}_i^* : i = 1, \cdots, L\}$ satisfying the required properties, we proceed as follows: The size of the code is $M' = 2^{m(I(U \wedge Y) - \epsilon)}$ and the common size of the subcodes is $M^* = 2^{m(I(U \wedge Z) + \gamma)}$, where the number γ is introduced in Lemma 1. This is possible because $I(U \wedge Y) > I(U \wedge Z)$. A subcode of this code can be selected by using a binary M'-sequence $s = (s_1, \cdots, s_{M'})$, where s_i is either 0 or 1. A codeword c_i^* is in the subcode $\mathcal{C}_s$ if and only if $s_i = 1$. After the code $\mathcal{C}_2$ is selected, we select $L = 2^{m\epsilon}$ subcodes of $\mathcal{C}_2$. Let these subcodes be $\mathcal{C}_i^*$. These codes are chosen randomly by letting for any s of weight M^*

$$\Pr(\mathcal{C}_i^* = \mathcal{C}_s) = \frac{1}{\binom{M'}{M^*}}$$

and the selections for different i are done independently. We are going to prove that for the random code chosen as above with probability approaching 1 as m goes to infinity the code has the required properties.

From Lemma 2, the subcode selected satisfies the following condition with probability approaching 1: for every $\mathcal{D} \subset \mathcal{Z}^m$, let $\tilde{Q}$ be the uniform distribution on the subcode, then

$$\tilde{Q}\tilde{V}^m(\mathcal{D}) \le (1+\epsilon)(1 - e^{-m\delta})^{-1} Q\tilde{V}^m(\mathcal{D}) + e^{-m\delta} \quad (42)$$
$$\tilde{Q}\tilde{V}^m(\mathcal{D}) \ge (1-\epsilon)(1 - e^{-m\delta}) Q\tilde{V}^m(\mathcal{D}) - e^{-m\delta}. \quad (43)$$

For sufficiently large m, we may assume that this probability is at most ϵ. We prove that if two subcode $\mathcal{C}_i$ and $\mathcal{C}_j$ both satisfy this condition, then

$$D(Q_i\tilde{V} \| Q_j\tilde{V}) \le \beta(\epsilon) \quad (44)$$

where $\beta(\epsilon)$ goes to zero as ϵ does and where Q_i denotes the uniform distribution on the code $\mathcal{C}_i^*$.

This is proved as follows: Let

$$\mathcal{D}_t = \left\{ z^m : \sum_{u^m} Q_i\tilde{V}^m(z^m \mid u^m) > t \sum_{u^m} Q_j\tilde{V}^m(z^m \mid u^m) \right\}$$

then

$$\begin{aligned} &(1+\epsilon)(1 - e^{-m\delta})^{-1} Q\tilde{V}^m(\mathcal{D}_t) + e^{-m\delta} \\ &\qquad > t(1-\epsilon)(1 - e^{-m\delta}) Q\tilde{V}^m(\mathcal{D}_t) - te^{-m\delta}. \end{aligned}$$

This implies

$$\begin{aligned} (1+t)e^{-m\delta} > &(t(1-\epsilon)(1 - e - m\delta) \\ &- (1+\epsilon)(1 - e^{-m\delta})^{-1}) Q\tilde{V}^m(\mathcal{D}_t) \end{aligned}$$

that is

$$Q\tilde{V}^m(\mathcal{D}_t) < \frac{(1+t)e^{-m\delta}}{(t(1-\epsilon)(1 - e^{-m\delta}) - (1+\epsilon)(1 - e^{-m\delta})^{-1})}.$$

We know that

$$\frac{\sum_{u^m} Q_i(u^m)\tilde{V}^m(z^m \mid u^m)}{\sum_{u^m} Q_j(u^m)\tilde{V}^m(z^m \mid u^m)}$$

is at most $e^{\alpha m}$, where

$$\alpha = \log \frac{\max \tilde{V}(z \mid u)}{\min \tilde{V}(z \mid u)}.$$

Letting

$$t = \frac{(1+\epsilon)^2 (1 - e^{-m\delta})^{-1}}{(1-\epsilon)(1 - e^{-m\delta})}$$

we obtain

$$D(Q_i\tilde{V} \| Q_j\tilde{V}) \le \frac{(1+t)e^{-m\delta}(1 - e^{-m\delta})}{\epsilon(1+\epsilon)} m\alpha + \log t. \quad (45)$$

We can easily see that the right-hand side of (45) approaches zero as m goes to infinity and then ϵ goes to zero.

By randomly selecting L subcodes, then deleting those subcodes for which the conditions in the Lemma are not satisfied, the number of remaining subcodes is in average at least $L(1-\epsilon)$. This is enough for our purpose.

We now prove that the intersection of two subcodes has more than ϵM^* codewords with doubly exponentially small probability. This is done by the following calculation:

$$\Pr\left(|\mathcal{C}_i \cap \mathcal{C}_j)| > \epsilon M^*\right) \leq M^* \frac{\binom{M^*}{\epsilon M^*}\binom{M'-M^*}{(1-\epsilon)M^*}}{\binom{M'}{M^*}}$$

which is doubly exponentially small. This proves that with probability approaching 1, any pair of the subcodes satisfy the condition that their intersection has a size at most ϵM^*. The proposition is proved.

D. Proof of the Converse Part

We begin with the following lemma.

Lemma 4: Let Q_1 and Q_2 be two distributions on $\mathcal{Z}^m$ and for any $\mathcal{V} \subset \mathcal{Z}^m$

$$Q_1(\mathcal{V}) + Q_2(\mathcal{V}^c) > 1 - \epsilon \tag{46}$$

and let U be a binary random variable with uniform distribution and $V(z^m \mid U = i) = Q_i$ for $i = 1, 2$ then

$$I(U \wedge Z^m) \leq \inf_{x>0} \frac{2}{x} + \log \frac{1}{1 - \frac{1}{2}x\epsilon}. \tag{47}$$

Proof: Let

$$\mathcal{V}_t = \{z^m : Q_1(z^m) < tQ_2(z^m)\}$$

then

$$Q_1(\mathcal{V}_1) + Q_2(\mathcal{V}_1^c) \leq tQ_2(\mathcal{V}_t) + 1 - Q_2(\mathcal{V}_t).$$

Therefore

$$tQ_2(\mathcal{V}_t) + 1 - Q_2(\mathcal{V}_t) > 1 - \epsilon.$$

This implies

$$(1-t)Q_2(\mathcal{V}_t) < \epsilon$$

that is, for $0 \leq t < 1$

$$Q_2(\mathcal{V}_t) < \frac{\epsilon}{1-t}.$$

Similarly, for $t > 1$

$$Q_1(\mathcal{V}_1^c) < \frac{t\epsilon}{t-1}.$$

Therefore, for any $0 \leq t < 1$

$$I(U \wedge Z^m) = \sum_{z^m} \frac{1}{2} Q_1(z^m) \log \frac{\frac{1}{2}Q_1(z^m)}{\frac{1}{4}(Q_1(z^m) + Q_2(z^m))} + \frac{1}{2} Q_2(z^m) \log \frac{\frac{1}{2}Q_2(z^m)}{\frac{1}{4}(Q_1(z^m) + Q_2(z^m))}.$$

For any t, $0 \leq t < 1$, we have for $z^m \notin \mathcal{V}_t \cup \mathcal{V}_{1/t}^c$

$$\left|\log \frac{\frac{1}{2}Q_1(z^m)}{\frac{1}{4}(Q_1(z^m) + q_2(z^m))}\right| \leq 1 + \log \frac{1}{1+t}$$

and for any z^m

$$\begin{aligned} \frac{1}{2} Q_1(z^m) \log &\frac{\frac{1}{2}Q_1(z^m)}{\frac{1}{4}(Q_1(z^m) + Q_2(z^m))} \\ &+ \frac{1}{2} Q_2(z^m) \log \frac{\frac{1}{2}Q_2(z^m)}{\frac{1}{4}(Q_1(z^m) + Q_2(z^m))} \\ &\leq \frac{1}{2}(Q_1(z^m) + Q_2(z^m)). \end{aligned}$$

Therefore

$$\begin{aligned} &I(U \wedge Z^m) \\ &\leq \sum_{z^m \in \mathcal{V}_t \cup \mathcal{V}_{1/t}^c} Q_1(z^m) \log \frac{\frac{1}{2}Q_1(z^m)}{\frac{1}{4}(Q_1(z^m) + Q_2(z^m))} \\ &\quad + \frac{1}{2} Q_2(z^m) \log \frac{\frac{1}{2}Q_2(z^m)}{\frac{1}{4}(Q_1(z^m) + Q_2(z^m))} \\ &\quad + 1 + \log \frac{1}{1+t} \\ &\leq \frac{1}{2}(Q_1(\mathcal{V}_t) + Q_2(\mathcal{V}_t) + Q_1(\mathcal{V}_{1/t}^c)) \\ &\quad + 1 + \log \frac{1}{1+t} \\ &\leq (1+t)\frac{\epsilon}{1-t} + 1 + \log \frac{1}{1+t}. \end{aligned}$$

Taking $t = 1 - x\epsilon$, we obtain

$$I(U \wedge Z^m) \leq \frac{2}{x} + \log \frac{1}{1 - \frac{1}{2}x\epsilon}.$$

This proves the lemma.

Lemma 5: Let Q_1 and Q_2 be two distributions on Z^m for which there exists a $\mathcal{V} \subset \mathcal{Z}^m$ such that

$$Q_1(\mathcal{V}) + Q_2(\mathcal{V}^c) < \epsilon \tag{48}$$

and let U be a binary random variable with uniform distribution and $V(z^m \mid U = i) = Q_i$ for $i = 1, 2$, then

$$I(U \wedge Z^m) \geq h\left(\frac{1}{2}(1-\epsilon)\right). \tag{49}$$

Proof:

$$\begin{aligned} I(U \wedge Z^m) \geq -\Bigg[&\frac{1}{2} Q_1(\mathcal{V}) \log \frac{Q_1(\mathcal{V}) + Q_2(\mathcal{V})}{2Q_1(\mathcal{V})} \\ &+ \frac{1}{2} Q_2(\mathcal{V}) \log \frac{Q_1(\mathcal{V}) + Q_2(\mathcal{V})}{2Q_2(\mathcal{V})} \\ &+ \frac{1}{2} Q_1(\mathcal{V}^c) \log \frac{Q_1(\mathcal{V}^c) + Q_2(\mathcal{V}^c)}{2Q_1(\mathcal{V}^c)} \\ &+ \frac{1}{2}(\mathcal{V}) \log \frac{Q_1(\mathcal{V}^c) + Q_2(\mathcal{V}^c)}{2Q_2(\mathcal{V}^c)}\Bigg] \\ \geq h\left(\frac{1}{2}(1-\epsilon)\right). \end{aligned} \tag{50}$$

The existence of the identification code at a positive rate implies the existence of distributions $Q(\cdot \mid i)$ and decoding regions $\mathcal{D}_i$ such that

1) for all i

$$\sum_{x^n \in \mathcal{X}^n} Q(x^n \mid i) W^n(\mathcal{D}_i \mid x^n) \geq 1 - \lambda. \tag{51}$$

2) for any pair $i \neq j$

$$\sum_{x^n \in \mathcal{X}^n} Q(x^n \mid j) W^n(\mathcal{D}_i \mid x^n) \leq \lambda \tag{52}$$

and

3) for any pair $i \neq j$ and any $\mathcal{V} \subset \mathcal{Z}^n$

$$\sum_{x^n \in \mathcal{X}^n} Q(x^n \mid j) V^n(\mathcal{V} \mid x^n) + \sum x^n \in \mathcal{X}^n Q(x^n \mid i) V^n(\mathcal{V}^c \mid x^n) \geq 1 - \lambda. \tag{53}$$

From the first two properties, we obtain

$$\sum_{x^n \in \mathcal{X}^n} Q(x^n \mid i) W^n(\mathcal{D}_i^c \mid x^n) + \sum_{x^n \in \mathcal{X}^n} Q(x^n \mid j) W^n(\mathcal{D}_i \mid x^n) \leq 2\lambda$$

which implies by Lemma 4 that for the random variable U defined there we have

$$I(U \wedge Y^n) \geq h\left(\frac{1}{2}(1 - 2\lambda)\right).$$

By the third property and Lemma 3, the same random variable U satisfies

$$I(U \wedge Z^n) \leq \inf_{x>0} \frac{2}{x} + \log \frac{1}{1 - \frac{1}{2}x\lambda}.$$

Then λ is small enough, we obtain the following conclusion: there exists a random variable U satisfying

i)

$$U \rightarrow X^m \rightarrow Y^m Z^m$$

ii)

$$I(U \wedge Y^m) > I(U \wedge Z^m).$$

Proposition 6: If there exists an m and a U satisfying the requirements i) and ii), then

$$C_s > 0.$$

Proof:

$$\begin{aligned} 0 < I(U \wedge Y^m) - I(U \wedge Z^m) &= \sum_{t=1}^{m} I(U \wedge Y_t \mid Y_1, \cdots, Y_{t-1}, Z_{t+1}, \cdots, Z_m) \\ &\quad - I(U \wedge Z_t \mid Y_1, \cdots, Y_{t-1}, Z_{t+1}, \cdots, Z_m). \end{aligned}$$

Therefore, there exists a t such that

$$\begin{aligned} &I(U \wedge Y_1 \mid Y_1, \cdots, Y_{t-1}, Z_{t+1}, \cdots, Z_m) \\ &\qquad - I(U \wedge Z_t \mid Y_1, \cdots, Y_{t-1}, Z_{t+1}, \cdots, Z_m) > 0 \end{aligned}$$

which implies

$$C_2 > 0.$$

The converse is proved.

Remark 4: Inspection of the proof shows that the possibility of "safe" identification for two options (or for 1 bit) implies "safe" identification at rate equal to Shannon's capacity.

References

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

[2] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* New York: Academic Press, 1981.

[3] R. Ahlswede and G. Dueck, "Identification via channels," *IEEE Trans. Inform. Theory*, vol. 35, pp. 15–29, 1989.

[4] ——, "Identification in the presence of feedback," *IEEE Trans. Inform. Theory*, vol. 35, pp. 30–39, 1989.

[5] T. S. Han and S. Verdú, "New results in the theory of identification via channels," *IEEE Trans. Inform. Theory*, vol. 38, no. 1, pp. 14–25, 1992.

[6] ——, "Approximation theory of output statistics," *IEEE Trans. Inform. Theory*, vol. 39, no. 3, 1993.

[7] R. Ahlswede and B. Verboven, "On identification via multiway channels with feedback," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1519–1526, 1991.

[8] R. Ahlswede, N. Cai, and Z. Zhang, "Erasure, list and detection zero error capacity for low noise and a relation to identification," SFB 343 "Diskrete Strukturen in der Mathematik," Univ. of Bielefeld, Preprint 93-068, 1993, submitted to IEEE.

[9] T. M. Cover and C. S. K. Leung, "An achievable rate region for the multiple access channel with feedback," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 292–298, 1981.

[10] R. Ahlswede, "A method of coding and an application to arbitrarily varying channels," *J. Comb. Inf. Syst. Sci.*, vol. 5, no. 1, pp. 10–35, 1980.