Uncertainty in Identification Systems

Minh Thanh Vu*, Tobias J. Oechtering*, Mikael Skoglund* and Holger Boche†
* Div. ISE, KTH Royal Institute of Technology
† LTI, Technische Universität München

Abstract

High-dimensional identification systems consisting of two groups of users in the presence of statistical uncertainties are considered in this work. The task is to design enrollment mappings to compress users' information and an identification mapping that combines the stored information in the database and an observation to estimate the underlying user index. The compression-identification trade-off regions are established for the compound, arbitrarily varying, general and mixture settings. It is shown that several settings admit the same compression-identification trade-offs. We then study a connection between the Wyner-Ahlswede-Körner network and the identification setting. It indicates that a strong converse for the WAK network is equivalent to a strong converse for the identification setting. Finally, we present strong converse arguments for the discrete identification setting that are extensible to the Gaussian scenario.

Index Terms

Identification systems, compound and arbitrarily varying, mixture distribution, information-spectrum method, strong converse, WAK-network.

I. Introduction

Identification systems have become ubiquitous in modern life with applications and aspects ranging from biometrics [1] to multimedia [2] to privacy-preserving identification [3]. Willems *et.al* [1] characterized the capacity of a biometric identification system with noisy data and observation sequences. The study has been developed further in several directions. In [4] the compression of users' information was taken into account. Therein, the author characterized the trade-off between the compression-identification rate for multi-stage identification systems. Independently, the compression of observation and users' information was also considered in [5] where achievable bounds for the compression-identification trade-off were proposed. Subsequently the work [6] extended [4] by including a distortion constraint on the reconstructed data sequence. Methods for improving search complexity and the corresponding limits were considered in [7], [8], and [9]. It is assumed in these works that users' information is iid generated from a known distribution and the observation channel is also perfectly known.

In this work we generalize the identification setting considered in [4]. We study an identification problem involving two groups of users. Each group uses their own compression mapping to enroll the corresponding users' data into a database. This setting is motivated from applications when separate databases are merged together for the identification purpose. We further assume that the distributions of users' data in two groups are distinct. As perfect knowledge of the distributions are usually unknown in practice and distribution learning is an active research theme in machine learning, we consider different models for data distribution in each group such as compound, arbitrarily varying, and mixture models. These models are standard in the source and channel coding literature [10]–[14]. We summarize our contributions in the following.

- We derive the optimal compression-identification trade-offs for the compound, arbitrarily varying, general and mixture settings. In [4] the true user's data sequence is identically independently distributed according to a known distribution and independent of the true index. This is not the case in the mentioned settings. The distribution of the chosen user's data is a non-iid mixture distribution and depends on the chosen user's index. Information spectrum arguments are used in combination with standard single-letterization arguments to show the converse in the considered settings.
- We generalize the connection between the identification setting and the Wyner-Ahlswede-Körner (WAK) network [15], [16]. Specifically, we show that given a WAK-code we can construct a corresponding identification scheme. Conversely, given an identification scheme there exists a corresponding WAK-code. The dual connection implies that a strong converse (an exponentially strong converse) for the WAK-setting is *equivalent* to a strong converse (an exponentially strong converse) for the identification problem. Hence, universally any approach that proves the strong converse for the WAK problem automatically proves the strong converse for the identification problem and vice versa.

The paper is organized as follows: The characterizations of the compound, arbitrary varying, and mixture settings are given in Section II, III and IV, respectively. In Section V we establish the duality between the WAK setting and the identification problem.

Notation: Random variables and their realizations are denoted by uppercase, and lowercase letters, respectively. The calligraphic letters are used to denote sets. X^n denotes the vector of random variables $(X_1, \dots, X_n)$ while x^n denotes its corresponding realization in the product set $\mathcal{X}^n$, unless otherwise stated. We also denote the probability measures by uppercase

This paper was presented in part at ISIT 2018.

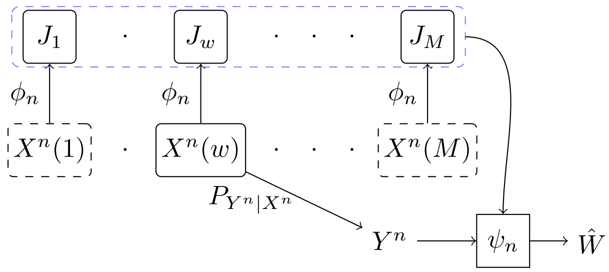

Fig. 1: The generic model for a single-group high-dimension identification system.

letters for example P_X while lowercase letters such as p_X are used for density functions, unless otherwise stated. For a set $\mathcal{A}$, $|\mathcal{A}|$ denotes its cardinality when $\mathcal{A}$ is finite whereas $\mathcal{A}^c$ denotes its complement. For two probability measures μ on $(\mathcal{A}, \mathcal{F}_{\mathcal{A}})$ and ν on $(\mathcal{B}, \mathcal{F}_{\mathcal{B}})$ we denote by $\mu \times \nu$ their product measure on $(\mathcal{A} \times \mathcal{B}, \mathcal{F}_{\mathcal{A}} \times \mathcal{F}_{\mathcal{B}})$. Given a probability measure μ on $\mathcal{A}$ and a Markov kernel $\kappa\colon \mathcal{A} \times \mathcal{F}_{\mathcal{B}} \to [0,1]$, $\kappa \times \mu$ denotes the joint probability measure on $(\mathcal{A} \times \mathcal{B}, \mathcal{F}_{\mathcal{A}} \times \mathcal{F}_{\mathcal{B}})$. Given a measure μ, $\mu^{\otimes n}$ denotes its n-fold product measure extension. $\log(\cdot)$ denotes the natural logarithm.

II. Compound setting

We first recap the generic setup for a single-group identification system with M users as depicted in Fig. 1. The identification process consists of two phases. In the *enrollment phase*, data sequences $(x^n(i))_{i=1}^M$ from M users are enrolled and stored as $(j_i)_{i=1}^M$ in a database via a compression mapping ϕ_n. It is assumed that the original data sequences $(x^n(i))_{i=1}^M$ might be absent after the enrollment phase is completed. In the *identification phase*, an observation y^n which is correlated with a selected sequence $x^n(w)$ is provided to the processing unit. The chosen index w is a realization of a uniform random variable W on $[1\colon M]$. The processing unit identifies the true user index w via an identification mapping ψ_n based on the observation y^n and the stored indices $(j_i)_{i=1}^M$.

In this section we consider a compound setting where users' data are generated from unknown distributions that belong to sets of distributions. Assume that there are two groups of users with M_1 and M_2 members, respectively. Let $M = M_1 + M_2$ be the total number of users in the system. We further assume that the fraction of users in the first group is given by $\lim_{n\to\infty} M_1/M = \alpha \in [0,1]$. Without loss of generality, we index the users in the first group by $[1\colon M_1]$ and the users in the second group by $[M_1+1\colon M]$.

Let

$$\mathcal{P}_1 = \{P_{X,s_1} \mid s_1 \in \mathcal{S}_1\},\ \mathcal{P}_2 = \{P_{X,s_2} \mid s_2 \in \mathcal{S}_2\}, \tag{1}$$

be two non-overlapping sets of probability measures defined on the same measurable space $(\mathcal{X}, \mathcal{F})$. Given an underlying state $s_1 \in \mathcal{S}_1$, the data sequences of the users in the first group are generated as $x^n(i) \sim P_{X,s_1}^{\otimes n}$ for $i \in [1\colon M_1]$. Similarly given an underlying state $s_2 \in \mathcal{S}_2$, the data sequences of the users in the second group are generated as $x^n(i) \sim P_{X,s_2}^{\otimes n}$ for $i \in [M_1+1\colon M]$. These assumptions represent the case that users in each group *all* belong to some communities and the distribution of each community is described by a state s_1 or s_2. An example for communities in the first and second group would be communities of users with the same fingerprint patterns and eye colors, respectively.

We define the overall set of data generating distributions $\mathcal{P}$ and the corresponding set of states $\mathcal{S}$ as

$$\mathcal{P} = \begin{cases} \mathcal{P}_1 & \text{if } \alpha = 1, \\ \mathcal{P}_1 \cup \mathcal{P}_2 & \text{if } \alpha \in (0,1), \\ \mathcal{P}_2 & \text{if } \alpha = 0, \end{cases} \quad \mathcal{S} = \begin{cases} \mathcal{S}_1 & \text{if } \alpha = 1, \\ \mathcal{S}_1 \cup \mathcal{S}_2 & \text{if } \alpha \in (0,1), \\ \mathcal{S}_2 & \text{if } \alpha = 0. \end{cases} \tag{2}$$

Furthermore, the observation channel is from the set

$$\mathcal{P}_c = \{P_{Y|X,\tau} \mid \tau \in \mathcal{T}\}. \tag{3}$$

The observation y^n is generated as follows. An index w is chosen uniformly at random from $[1\colon M]$. For a given τ, y^n is observed through the channel $P_{Y|X,\tau}^{\otimes}$ with input $x^n(w)$. An example for the set $\mathcal{P}_c$ could be different deformities for biometric patterns.

Throughout the paper we assume that $\mathcal{S}_1$, $\mathcal{S}_2$, and $\mathcal{T}$ are finite, unless otherwise stated. Furthermore, we denote the corresponding set of marginal distributions on $\mathcal{Y}$ by

$$\mathcal{P}_{\mathcal{Y}} = \{P_Y \mid P_Y \text{ is the marginal distribution on } \mathcal{Y} \text{ of } P_{Y|X,\tau} \times P_{X,s},\ \text{for some } (\tau, s) \in \mathcal{T} \times \mathcal{S}\}. \tag{4}$$

For simplicity, we enumerate elements of $\mathcal{P}_{\mathcal{Y}}$ by $P_{Y,\kappa}$ where $\kappa \in [1 : |\mathcal{P}_{\mathcal{Y}}|]$.

Definition 1. *A identification scheme consists of two group enrollment mappings*

$$\phi_{kn}\colon \mathcal{X}^n \to \mathcal{M}_k,\ k=1,2, \tag{5}$$

which compress the corresponding users' information and store it in a database, and an identification mapping

$$\psi_n\colon \mathcal{Y}^n \times \mathcal{M}_1^{M_1} \times \mathcal{M}_2^{M_2} \to [1:M] \cup \{e\}, \tag{6}$$

which identifies the true user using the observation and the stored information in the database.

For a given triple (s_1, s_2, τ) the corresponding probability of error is given as

$$\begin{aligned}\Pr_{s_1,s_2,\tau}\{W \neq \hat{W}\} = \sum_w \frac{1}{M} \int dP^{\otimes n}_{Y|X,\tau}(y^n|x^n(w)) \prod_{i=1}^{M_1} dP^{\otimes n}_{X,s_1}(x^n(i)) \prod_{i=M_1+1}^{M} dP^{\otimes n}_{X,s_2}(x^n(i))\\ \mathbf{1}\{w \neq \psi_n(y^n, (\phi_{1n}(x^n(i)))_{i=1}^{M_1}, (\phi_{2n}(x^n(i)))_{i=M_1+1}^{M})\}.\end{aligned} \tag{7}$$

Definition 2. *For a given $\alpha \in [0,1]$, a compression-identification rate pair (R_c, R_i) is achievable if there exists a triple of mappings $(\phi_{1n}, \phi_{2n}, \psi_n)$ such that*

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n}\log|\mathcal{M}_k| \leq R_c,\ \forall k=1,2, \quad \liminf_{n\to\infty}\frac{1}{n}\log M \geq R_i,\\ \lim_{n\to\infty} \sup_{(s_1,s_2,\tau)\in\mathcal{S}_1\times\mathcal{S}_2\times\mathcal{T}} \Pr_{s_1,s_2,\tau}\{W\neq\hat{W}\} = 0.\end{aligned} \tag{8}$$

The closure of the set of all achievable rate pair is denoted by $\mathcal{R}_{\rm sc}$.

Definition 3. *When the alphabets $\mathcal{X}$ and $\mathcal{Y}$ are finite, we define $\bar{\mathcal{R}}_{\rm sc}$ to be the set of rate pairs (R_c, R_i) for which we have*

$$R_c \geq \max_{s\in\mathcal{S}} I(X_s; U_s), \quad R_i \leq \min_{(s,\tau)\in\mathcal{S}\times\mathcal{T}} I(Y_\tau; U_s), \tag{9}$$

where $\mathcal{S}$ is defined as in (2), *$P_{Y_\tau X_s U_s} = P_{Y|X,\tau} \times P_{X,s} \times P_{U|X,s}$, and U_s is defined on an alphabet $\mathcal{U}$ with $|\mathcal{U}| \leq |\mathcal{X}| + |\mathcal{T}|$.*

If $\mathcal{X}$ and $\mathcal{Y}$ are $\mathbb{R}$, $\mathcal{P}$ can be represented by a set of density functions $p_{X,s}$ and $\mathcal{P}_c$ can be represented by the set of conditional density functions $p_{Y|X,\tau}$ then $\bar{\mathcal{R}}_{\rm sc}$ is defined similarly as the closure of rate pairs (R_c, R_i) satisfying (9) *over the set of test channels*[1] *$p_{U|X,s}$.*

Theorem 1. *For a given fraction of the first group $\alpha \in [0,1]$ as well as given sets $\mathcal{P}_1$, $\mathcal{P}_2$ and $\mathcal{P}_c$, the region $\bar{\mathcal{R}}_{\rm sc}$ is achievable in the mentioned cases, i.e.,*

$$\bar{\mathcal{R}}_{\rm sc} \subseteq \mathcal{R}_{\rm sc}. \tag{10}$$

Furthermore, if further $\mathcal{X}$ and $\mathcal{Y}$ are finite then we have $\mathcal{R}_{\rm sc} = \bar{\mathcal{R}}_{\rm sc}$.

The achievability proof of Theorem 1 follows the one of Theorem 2 in Section III. The converse proof of Theorem 1 is given in Appendix B. Our result generalizes naturally to settings with a finite number of groups. However for clarity we present the setting and state the corresponding result with only two groups.

Remark 1. *When $M_1 = M$ and $|\mathcal{S}_1| = 1$ or $M_2 = M$ and $|\mathcal{S}_2| = 1$ hold, our result reduces to the one given in [4, Theorem 1]. Note that when $1 < M_1 < M$ the distribution of the data sequence of the chosen user $X^n(W)$ is given by*

$$P_{X^n(W)} = \frac{M_1}{M} P^{\otimes n}_{X,s_1} + \left(1 - \frac{M_1}{M}\right) P^{\otimes n}_{X,s_2}. \tag{11}$$

Hence $X^n(W)$ is not iid and neither $X^n(W) \sim P^{\otimes n}_{X,s_1}$ nor $X^n(W) \sim P^{\otimes n}_{X,s_2}$ holds. $X^n(W)$ is also not independent of W as in [4]. This further leads to Y^n be neither iid nor independent of W. Therefore the standard converse steps used in [4] are not straightforwardly applicable. To show the converse of Theorem 1 we use the information-spectrum approach in combination with standard single-letterization steps. The idea of the proof is to show that the two-group identification setting can be decomposed into two sub-identification problems corresponding to the first and second groups, respectively. Inside each group $(Y^n, X^n(W))$ is jointly iid given the states. This property is summarized in Lemma 2 in Appendix A. When $\alpha \in (0,1)$ it can be seen that $\frac{1}{n}\log M_k \approx \frac{1}{n}\log M$ holds for $k=1,2$. Therefore, identification rate constraints in the two sub-problems become relevant and independent of the specific value of α.

[1] We require that the corresponding entropy and mutual information are finite.

Example: Assume that $M_1 = M$. Let X be a zero mean Gaussian random variable with unknown variance which belongs to the set $\{\sigma_1^2, \sigma_2^2\}$. Without loss of generality we assume that $0 < \sigma_1^2 < \sigma_2^2$. It is clear that for any probability metric $d(\cdot,\cdot)$, we have $d(P_1, P_2) \neq 0$. The AWGN observation channel is modeled by

$$Y = X + N, \quad N \sim \mathcal{N}(0, \sigma_N^2). \tag{12}$$

It can be seen that the compression-identification rate region $\mathcal{R}_{\mathrm{sc}}$ is given by

$$\begin{aligned} R_c &\geq \frac{1}{2}\log\frac{\sigma_1^2}{\sigma_{Y_1}^2 e^{-2R_i} - \sigma_N^2} \\ 0 \leq R_i &< \frac{1}{2}\log(1 + \sigma_1^2/\sigma_N^2), \end{aligned} \tag{13}$$

where $\sigma_{Y_1}^2 = \sigma_1^2 + \sigma_N^2$. The achievability follows from Theorem 1 with *test* channels $p_{U|X,1}$ and $p_{U|X,2}$ such that given each state we can write

$$X = U + N', \tag{14}$$

where U and N' are independent Gaussian rvs. More specifically, the distribution of U given the first state is given by $P_{U,1} = \mathcal{N}\big(0, \sigma_{Y_1}^2(1 - e^{-2R_i})\big)$. Similarly we have $P_{U,2} = \mathcal{N}\big(0, \sigma_{Y_2}^2(1 - e^{-2R_i})\big)$. The converse holds due to the entropy power inequality. The example easily illustrates that our scheme needs to adapt to the worst state.

III. Arbitrarily varying setting

In this section we consider the scenario where each user $i \in [1:M]$ has its own state. Similar to Section II we assume that there exists two groups of users with the corresponding numbers of users M_1 and M_2, respectively. The fraction of users in the first group is similarly denoted by α. For each user in the first group the corresponding data sequence $x^n(i)$ is generated independently from $P_{X,s_i}^{\otimes n}$ where $s_i \in \mathcal{P}_1$ for all $i \in [1:M_1]$. The data sequence of each user in the second group is generated as $x^n(i) \sim P_{x,\tilde{s}_i}^{\otimes n}$ where $\tilde{s}_i \in \mathcal{S}_2$ for all $i \in [M_1+1:M]$. Compared to the setting of Theorem 1, the underlying states can be different from user to user. An example of this setting would be the fingerprints (eye colors) of users in the first (second) group do not follow the same patterns.

We define the set of overall source distributions as in (2). The observation channel is similarly assumed to be in the set $\mathcal{P}_c$ as in (3). The definition of an identification scheme is identical as the one given in Definition 1. For simplicity we denote by $\boldsymbol{s}$ a sequence of states $((s_i)_{i=1}^{M_1}, (\tilde{s}_i)_{i=M_1+1}^{M}) \in \mathcal{S}_1^{M_1} \times \mathcal{S}_2^{M_2}$. The corresponding probability of error is given by

$$\begin{aligned} \Pr_{\boldsymbol{s},\tau}\{W \neq \hat{W}\} = \sum_w \frac{1}{M} \int dP_{Y|X,\tau}^{\otimes n}(y^n|x^n(w)) \prod_{i=1}^{M_1} dP_{X,s_i}^{\otimes n}(x^n(i)) \prod_{i=M_1+1}^{M} dP_{X,\tilde{s}_i}^{\otimes n}(x^n(i)) \\ \mathbf{1}\{w \neq \psi_n(y^n, (\phi_{1n}(x^n(i)))_{i=1}^{M_1}, (\phi_{2n}(x^n(i)))_{i=M_1+1}^{M})\}. \end{aligned} \tag{15}$$

Similarly, we have the following definition of achievability.

Definition 4. *For a given $\alpha \in [0,1]$, a compression-identification pair (R_c, R_i) is achievable if there exists a triple of mappings $(\phi_{1n}, \phi_{2n}, \psi_n)$ such that*

$$\begin{aligned} &\limsup_{n\to\infty} \frac{1}{n}\log|\mathcal{M}_k| \leq R_c, \ \forall k = 1,2, \quad \liminf_{n\to\infty} \frac{1}{n}\log M \geq R_i, \\ &\lim_{n\to\infty} \sup_{\boldsymbol{s}\in\mathcal{S}_1^{M_1}\times\mathcal{S}_2^{M_2}, \tau\in\mathcal{T}} \Pr_{\boldsymbol{s},\tau}\{W \neq \hat{W}\} = 0. \end{aligned} \tag{16}$$

The closure of the set of all achievable pairs is denoted by $\mathcal{R}_{\mathrm{iis}}$.

Note that the number of constraints in the current setting grows exponentially with the block length n. It can be seen that when $\boldsymbol{s} = (\underbrace{s_1,\ldots,s_1}_{M_1}, \underbrace{s_2,\ldots,s_2}_{M_2})$ where $s_1 \in \mathcal{S}_1$ and $s_2 \in \mathcal{S}_2$ the error probability given in (15) equals to the one given in (7) provided that the same triple of mappings are used in both settings. Therefore, $\mathcal{R}_{\mathrm{iss}} \subseteq \mathcal{R}_{\mathrm{sc}}$. The result for the current setting is summarized in the following theorem.

Theorem 2. *For given set of channels $\mathcal{P}_c$ and sets of distributions $\mathcal{P}_1$, $\mathcal{P}_2$, we have*

$$\bar{\mathcal{R}}_{\mathrm{sc}} \subseteq \mathcal{R}_{\mathrm{iis}}. \tag{17}$$

When $\mathcal{X}$ and $\mathcal{Y}$ are finite, we further have $\mathcal{R}_{\mathrm{iss}} = \bar{\mathcal{R}}_{\mathrm{sc}}$.

The achievability proof of Theorem 2 is given in the Appendix C while the converse proof of Theorem 2 follows the one of Theorem 1. Similarly as in Remark 1 given an arbitrary but unknown sequence of states $\boldsymbol{s}$ the data sequence of the chosen user $X^n(W)$ is likewise not iid and independent of W. In the following remark we present the proof idea of Theorem 2.

Remark 2. *For simplicity we consider the case $\alpha \in (0,1)$. Since given an arbitrary but unknown sequence of states $\boldsymbol{s}$, the data sequence of each user is iid generated according to either $P_{X,s_1}^{\otimes n}$, $s_1 \in \mathcal{S}_1$ or $P_{X,s_2}^{\otimes n}$, $s_2 \in \mathcal{S}_2$ we can use the concentration properties of probability measures to estimate the unknown state. Given an estimate $\hat{s}(i)$ for the i-th user, we use a corresponding codebook generated from $P_{U,\hat{s}(i)}^{\otimes n}$ to compress the data sequence $x^n(i)$ and store the index $m_{i,\hat{s}(i)}$ inside the database. Given an arbitrary but unknown channel state τ we also can estimate the state of the output distribution of y^n denoted by $\hat{\kappa}$. Given the estimate $\hat{\kappa}$ we find all the pairs of source and channel states $(s,\tau) \in \mathcal{S} \times \mathcal{T}$ such that $P_{Y,\hat{\kappa}}$ is the marginal of $P_{Y|X,\tau} \times P_{X,s}$. In the identification process, we look for a unique index $\hat{w}$ such that $(y^n, u^n(m_{\hat{w},\hat{s}(\hat{w})}))$ belongs to the union of the joint typical sets $\mathcal{A}_\gamma^n(P_{Y_\tau U_s})$ over all such pairs (s,τ). We note that in this problem we do not need to store the estimated states $\hat{s}(i)$ along with the compressed index $m_{i,\hat{s}(i)}$ inside the database. Note further that inside each group the estimated state $\hat{s}(i)$ can be different from user to user. The only requirement is that the estimated state of the true user is correct with high probability.*
The case $\alpha = 0$ or $\alpha = 1$ can be resolved by controlling the error probability of users from the major group as the error probability of users from the minor group contributes only a vanishing fraction to the total error.

IV. Mixture models

In this section we continue studying the two group identification models mentioned in Section II and Section III. We again assume M_1 and M_2 users for two groups and the same sets $\mathcal{P}_1$, $\mathcal{P}_2$, and $\mathcal{P}_c$. However, in contrast to the previous models we assume in this section that the data sequences of users in group k are generated independently from distribution $P_{X_k^n}$ given in the following

$$P_{X_k^n}(x^n) = \sum_{s \in \mathcal{S}_k} \alpha_{ks} P_{X,s}^{\otimes n}(x^n),\ k = 1,2, \tag{18}$$

where $(\alpha_{ks})_{s \in \mathcal{S}_k}$ are arbitrary but fixed tuples which satisfy $\sum_{s \in \mathcal{S}_k} \alpha_{ks} = 1$, and $\alpha_{ks} > 0$, for all $k = 1,2$ and $s \in \mathcal{S}_k$. We assume further that the observation channel is given as

$$P_{Y^n|X^n} = \sum_{\tau \in \mathcal{T}} \alpha_\tau P_{Y|X,\tau}^{\otimes n}, \tag{19}$$

where $\mathcal{T}$ is a finite set, $\sum_{\tau \in \mathcal{T}} \alpha_\tau = 1$ and $\alpha_\tau > 0$ for all $\tau \in \mathcal{T}$. In this setting we know imperfectly that data of users in the the k-th group are independently generated from $P_{X,s}^{\otimes n}$ with probability α_{ks} for $k = 1,2$, $s \in \mathcal{S}_k$ and the observation channel is $P_{Y|X,\tau}^{\otimes n}$ with probability α_τ for $\tau \in \mathcal{T}$. Additionally, we consider only finite alphabets $\mathcal{X}$ and $\mathcal{Y}$ in this subsection.

A. A general identification-compression trade-off

Before providing the detailed characterization of the region for the motivating setting, we make a digression and consider a more general problem where the underlying processes are not necessarily memoryless. Assume that data of the users in the first and second groups are generated independently from the distributions $P_{X_1^n}$ and $P_{X_2^n}$ defined on a finite alphabet $\mathcal{X}_n$, respectively. $P_{X_1^n}$ and $P_{X_2^n}$ are not necessary distributions of memoryless sources. The data sequences are enrolled inside a database by two compression mappings ϕ_{1n} and ϕ_{2n}. The observation channel is given by $P_{Y^n|X^n}$ which is not necessary a discrete memoryless channel. To study this general problem we use the following definition of achievability.

Definition 5. *For a given $\alpha \in [0,1]$, a rate pair (R_c, R_i) is achievable for the general identification problem if there exists a pair of identification mappings (ϕ_{1n}, ϕ_{2n}) and an identification mapping ψ_n such that*

$$\begin{aligned} \limsup_{n\to\infty} \frac{1}{n} \log|\mathcal{M}_k| \le R_c,\ \forall k = 1,2,\ \liminf_{n\to\infty} \frac{1}{n}\log M \ge R_i, \\ \lim_{n\to\infty} \Pr\{\hat{W} \ne W\} = 0. \end{aligned} \tag{20}$$

Let $\mathcal{R}_{\rm gen}$ denote the closure of the set of all achievable rate pairs (R_c, R_i).

To characterize $\mathcal{R}_{\rm gen}$ we need the following quantities. For a joint discrete process $(\mathbf{X}, \mathbf{Y}) = \{(X^n, Y^n)\}_{n=1}^\infty$ such that $(X^n, Y^n) \sim P_{X^n Y^n}$, the spectral sup-mutual information (inf-mutual information) rate [14] is defined respectively as

$$\begin{aligned} \bar{I}(\mathbf{X};\mathbf{Y}) &= \text{p-}\limsup_{\text{n}\to\infty} \frac{1}{n}\log\frac{P_{Y^n|X^n}(Y^n|X^n)}{P_{Y^n}(Y^n)} = \inf\left\{\beta \Big| \lim_{n\to\infty} \Pr\left[\frac{1}{n}\log\frac{P_{Y^n|X^n}(Y^n|X^n)}{P_{Y^n}(Y^n)} > \beta\right] = 0\right\}, \\ \underline{I}(\mathbf{X};\mathbf{Y}) &= \text{p-}\liminf_{\text{n}\to\infty} \frac{1}{n}\log\frac{P_{Y^n|X^n}(Y^n|X^n)}{P_{Y^n}(Y^n)} = \sup\left\{\beta \Big| \lim_{n\to\infty} \Pr\left[\frac{1}{n}\log\frac{P_{Y^n|X^n}(Y^n|X^n)}{P_{Y^n}(Y^n)} < \beta\right] = 0\right\}. \end{aligned} \tag{21}$$

Theorem 3. *When $\alpha \in (0,1)$, $\mathcal{R}_{\rm gen}$ is the closure of the set of all pairs (R_c, R_i) such that*

$$\begin{aligned} R_c &\ge \max\{\bar{I}(\mathbf{X}_1;\mathbf{U}_1), \bar{I}(\mathbf{X}_2;\mathbf{U}_2)\} \\ R_i &\le \min\{\underline{I}(\mathbf{Y}_1;\mathbf{U}_1), \underline{I}(\mathbf{Y}_2;\mathbf{U}_2)\}, \end{aligned} \tag{22}$$

for some general processes $\mathbf{U}_1 = \{U_1^n\}_{n=1}^\infty$ *and* $\mathbf{U}_2 = \{U_2^n\}_{n=1}^\infty$*.* U_k^n *is defined on a finite alphabet* $\mathcal{U}_{kn}$ *such that there exists* $N_0 > 0$ *for which* $\frac{1}{n}\log|\mathcal{U}_{kn}| < N_0$*,* $\forall n$*,* $\forall k = 1, 2$*. Additionally, for each* n *and* k*, the joint distribution satisfies* $P_{Y_k^n X_k^n U_k^n} = P_{Y^n|X^n} \times P_{X_k^n U_k^n}$*. When* $\alpha = 1$ *(*$\alpha = 0$*) holds the second (first) terms inside the maximization-minimization in* (22) *can be omitted.*

Remark 3. *In the proof of Theorem 2 we use the union of jointly typical sets over all compatible source and channel states as the decoding set. The arguments hold since given the states and chosen user, the distribution of the observation sequence is iid. It is not straightforward to use similar arguments to the general case where given the states and the chosen user the distribution of the observation sequence is not necessary iid. We instead use the "typical set" of the following mixed distribution*

$$P_{\tilde{Y}^n\tilde{U}^n} = \alpha P_{Y_1^n U_1^n} + (1-\alpha) P_{Y_2^n U_2^n}. \tag{23}$$

The proof of Theorem 3 is presented in Appendix D.

B. Generalized mixture models

We now turn back to the motivating setting at the beginning of this section. We use Definition 5 as the definition of achievability for this setting and denote the corresponding closure of set of achievable rate pairs by $\mathcal{R}_{\text{mix}}$.
For a given $\alpha \in [0, 1]$, block length n, let $(\phi_{1n}, \phi_{2n}, \psi_n)$ be an identification scheme for the arbitrarily varying setting in Section III. If the triple of mappings is applied to the current mixture model given in (18) and (19), we then observe that

$$\Pr\{\hat{W} \neq W\} \leq \sup_{\boldsymbol{s}\in\mathcal{S}_1^{M_1}\times\mathcal{S}_2^{M_2}, \tau\in\mathcal{T}} \Pr_{\boldsymbol{s},\tau}\{\hat{W} \neq W\}. \tag{24}$$

The left-hand side of the above inequality is the error probability in the current mixture model while the right-hand side is the supremum of identification error probabilities (15) in the arbitrarily varying setting in Section III. This implies that if a pair (R_c, R_i) is achievable for the arbitrarily varying setting then it is also achievable for the mixture model. Therefore we have $\mathcal{R}_{\text{iss}} \subseteq \mathcal{R}_{\text{mix}}$. The following theorem provides the characterization for $\mathcal{R}_{\text{mix}}$.

Theorem 4. *Assume that* $\mathcal{X}$ *and* $\mathcal{Y}$ *are finite. For a given* $\alpha \in [0, 1]$*, we have* $\mathcal{R}_{\text{mix}} = \bar{\mathcal{R}}_{\text{sc}}$*.*

It can be seen from Theorem 1, 2 and 4 that the regions $\mathcal{R}_{\text{sc}}$, $\mathcal{R}_{\text{iss}}$ and $\mathcal{R}_{\text{mix}}$ have the same structure when $\mathcal{X}$ and $\mathcal{Y}$ are finite.

Proof: As $\mathcal{X}$ and $\mathcal{Y}$ are finite we have $\mathcal{R}_{\text{iis}} = \bar{\mathcal{R}}_{\text{sc}}$ by Theorem 2. Therefore $\bar{\mathcal{R}}_{\text{sc}} \subseteq \mathcal{R}_{\text{mix}}$ holds. Assume that there exists an identification scheme such that the rate pair (R_c, R_i) is achievable when $\alpha \in (0, 1)$. Given an arbitrary $\gamma > 0$ for sufficiently large n we have $\log M \geq n(R_i - \gamma)$. Applying Lemma 2 with $P_{X_1^n}$, $P_{X_2^n}$, and $P_{Y^n|X^n}$, we have

$$R_i \leq \min\{\underline{I}(\mathbf{Y}_k; \phi_k(\mathbf{X}_k))\}_{k=1}^2, \tag{25}$$

where $(\mathbf{Y}_k, \phi_k(\mathbf{X}_k)) = \{(Y_k^n, \phi_{kn}(X_k^n))\}_{k=1}^\infty$, $(Y_k^n, \phi_{kn}(X_k^n)) \sim P_{Y_k^n \phi_{kn}(X_k^n)}$ and

$$P_{Y_k^n \phi_{kn}(X_k^n)}(y^n, m_k) = \sum_{\tau\in\mathcal{T}, s_k\in\mathcal{S}_k} \alpha_\tau \alpha_{k s_k} \sum_{x^n:\ \phi_{kn}(x^n) = m_k} P_{Y|X,\tau}^{\otimes n} \times P_{X,s_k}^{\otimes n}(y^n, x^n). \tag{26}$$

Therefore we have,

$$R_i \leq \underline{I}(\mathbf{Y}_\tau; \phi_k(\mathbf{X}_{s_k})),\ \forall k = 1, 2,\ \forall (s_k, \tau) \in \mathcal{S}_k \times \mathcal{T}, \tag{27}$$

which is exactly (66) in the proof of Theorem 1. The rest of the proof therefore follows the one of Theorem 1. ■

In the rest of this section we consider another mixture model. We assume for simplicity that there is only a single group in our system, i.e., $M_1 = M$. Furthermore we assume that the observation channel is given by (19). The distribution of data sequence is given by

$$P_{X^n} = \sum_{s\in\mathcal{S}} \alpha_s P_{X,s}^{\otimes n} \tag{28}$$

where $\mathcal{S}$ is a *countably infinite* alphabet, $\alpha_s > 0$ for all $s \in \mathcal{S}$ and $\sum_{s\in\mathcal{S}} \alpha_s = 1$. The definition of achievability is similarly given as in Definition 5. Denote the closure of the set of all achievable rate pairs by $\mathcal{R}_{\text{enum}}$ and define $R_{\max}^{\text{enum}}(R_c) = \sup\{R_i | (R_c, R_i) \in \mathcal{R}_{\text{enum}}\}$. The following theorem characterizes $R_{\max}^{\text{enum}}(R_c)$ for the setting given by (28).

Theorem 5. *When* $\mathcal{X}$ *and* $\mathcal{Y}$ *are finite, we have*

$$R_{\max}^{\text{enum}}(R_c) = \inf_{s\in\mathcal{S}} \theta^s(R_c), \tag{29}$$

where $\theta^s(R_c)$ *is defined as*

$$\theta^s(R_c) = \max_{\substack{P_{U|X,s}:\ |\mathcal{U}|\leq|\mathcal{X}|+|\mathcal{T}| \\ I(X_s;U_s)\leq R_c}} \min_{\tau\in\mathcal{T}} I(Y_\tau; U_s),\ \text{where } P_{Y_\tau X_s U_s} = P_{Y|X,\tau} \times P_{U_s X_s},\ \forall \tau \in \mathcal{T}. \tag{30}$$

The proof of Theorem 5 is given in Appendix E.

V. Duality between Wyner-Ahlswede-Körner network and the identification problem

We establish herein a connection between the identification problem and the problem of lossless source coding with coded side information, which sheds some light on reasons why the expression (17) holds. We consider the single group setup with known distribution and channel in the identification problem. Recall that a code for Wyner-Ahlswede-Körner network, called a WAK-code, [15], [16] for the pair of discrete memoryless sources $(\bar{X}^n, \bar{Y}^n) \sim P_{XY}^{\otimes n}$ consists of three mappings,

$$\begin{aligned} \phi_{1n}\colon \mathcal{X}^n \to \mathcal{M}_1, \quad \phi_{2n}\colon \mathcal{Y}^n \to \mathcal{M}_2, \\ \psi_n\colon \mathcal{M}_1 \times \mathcal{M}_2 \to \mathcal{Y}^n. \end{aligned} \tag{31}$$

Definition 6. *Given an $\epsilon \in [0,1)$, a pair (R_1, R_2) is ϵ-achievable if there exists an WAK-code such that*

$$\begin{aligned} \limsup_{n\to\infty} \frac{1}{n} \log |\mathcal{M}_k| \leq R_k,\ k = 1,2, \\ \limsup_{n\to\infty} \Pr\{Y^n \neq \hat{Y}^n\} \leq \epsilon. \end{aligned} \tag{32}$$

The closure of the set of all achievable rate pairs is denoted by $\mathcal{R}_{\mathrm{WAK},\epsilon}$.

It is well-known that the region $\mathcal{R}_{\mathrm{WAK},\epsilon}$ is characterized by the following conditions

$$\begin{aligned} R_1 \geq I(X;U),\ R_2 \geq H(Y|U), \\ U - X - Y, \quad |\mathcal{U}| \leq |\mathcal{X}| + 1,\ \forall \epsilon \in [0,1). \end{aligned} \tag{33}$$

For a single-group identification system with $X^n(i) \sim P_X^{\otimes n}$ for all $i \in [1:M]$ and the observation channel $P_{Y|X}^{\otimes n}$ the *ϵ-achievable* definition is given in the following.

Definition 7. *Given an $\epsilon \in [0,1)$, a pair (R_c, R_i) is ϵ-achievable if there exists an identification scheme such that*

$$\begin{aligned} \limsup_{n\to\infty} \frac{1}{n} \log |\mathcal{M}_1| \leq R_c,\ \liminf_{n\to\infty} \frac{1}{n} \log M \geq R_i, \\ \limsup_{n\to\infty} \Pr\{\hat{W} \neq W\} \leq \epsilon. \end{aligned} \tag{34}$$

The closure of the set of all achievable rate pairs is denoted by $\mathcal{R}_{\mathrm{ID},\epsilon}$.

We first have the following observation which is a strong converse w.r.t. the identification rate, see also [17] for another strong converse proof using Arimoto's argument.

Proposition 1. *Given $\epsilon \in [0,1)$, if $(R_c, R_i) \in \mathcal{R}_{\mathrm{ID},\epsilon}$ then $R_i \leq I(Y;X)$.*

Proof. Suppose that $(R_c, R_i) \in \mathcal{R}_{\mathrm{ID},\epsilon}$ with $R_i = I(X;Y) + 3\gamma$ for some $\gamma > 0$. It suffices to consider the uncompressed scenario. Applying Lemma 2 with $M_1 = M$, $P_{X_1^n} = P_X^{\otimes n}$ and $P_{Y^n|X^n} = P_{Y|X}^{\otimes n}$ as well as using the weak law of large numbers we obtain

$$\Pr\{\hat{W} \neq W\} \to 1, \text{ as } n \to \infty, \tag{35}$$

for any identification mapping ψ_n, a contradiction to $(R_c, R_i) \in \mathcal{R}_{\mathrm{ID},\epsilon}$. □

In [18, Section III.C and Section III.E] the authors stated the following relation between the WAK problem and the identification problem for $\epsilon = 0$,

$$\mathcal{R}_{\mathrm{ID},0} = \{(R_1, H(Y) - R_2) \mid (R_1, R_2) \in \mathcal{R}_{\mathrm{WAK},0}\}. \tag{36}$$

In both problems R_i characterizes the number of confused (coded) codewords that the system can tolerate. The authors also stated briefly that the entropy characterization approach in their work provides the optimal achievable region for the identification problem.

In the following we relate both problem using a generalization of the entropy characterization argument. For an arbitrary but given $\epsilon \in [0,1)$, we show that given an *arbitrarily* ϵ-achievable code for the WAK-setting we can construct a corresponding ϵ-achievable scheme for the identification setting with a corresponding rate pair. Conversely, given any ϵ-achievable identification scheme, we show that there exists a corresponding ϵ-achievable WAK-code.

Theorem 6. *Given $\epsilon \in [0,1)$.*

1) *Assume that there exists a WAK-code $(\phi_{1n}, \phi_{2n}, \psi_n)$ such that $(R_1, R_2) \in \mathcal{R}_{\mathrm{WAK},\epsilon}$ is ϵ-achievable, where $R_2 < H(Y)$ then we can construct an identification scheme (ϕ_{1n}, ψ'_n) based on the given* WAK*-code such that $(R_1, H(Y) - R_2)$ is in $\mathcal{R}_{\mathrm{ID},\epsilon}$.*
2) *Assume that there exists an identification scheme (ϕ_n, ψ_n) such that $(R_i, R_c) \in \mathcal{R}_{\mathrm{ID},\epsilon}$ is ϵ-achievable where $R_i < H(Y)$. Then there exists a* WAK*-code $(\phi_n, \phi'_{2n}, \psi'_n)$ corresponding to the provided identification scheme such that $(R_c, H(Y) - R_i)$ is in $\mathcal{R}_{\mathrm{WAK},\epsilon}$.*

Consequently, given $(R_a, R_b) \in \mathbb{R}_+^2$ with $R_b \le H(Y)$,

$$(R_a, R_b) \in \mathcal{R}_{\mathrm{WAK},\epsilon} \Leftrightarrow (R_a, H(Y) - R_b) \in \mathcal{R}_{\mathrm{ID},\epsilon}, \quad \forall \epsilon \in [0,1). \tag{37}$$

Remark 4. *Theorem 6 and Proposition 1 imply that for each $\epsilon > 0$, $\mathcal{R}_{\mathrm{ID},\epsilon}$ corresponds to the sub-region[2] of $\mathcal{R}_{\mathrm{WAK},\epsilon}$ with $R_2 \le H(Y)$. Hence a strong converse for the WAK-problem is equivalent to a strong converse for the identification problem. In fact our arguments also indicate that an exponentially strong converse for the WAK-problem is equivalent to an exponentially strong converse for the identification problem.*

Proof: WAK $\Rightarrow$ ID: Fix an arbitrary $\gamma > 0$. Suppose that $(\phi_{1n}, \phi_{2n}, \psi_n)$ is a WAK-code for which $(R_1, R_2) \in \mathcal{R}_{\mathrm{WAK},\epsilon}$ is ϵ-achievable. We define for each $m_1 \in \mathcal{M}_1$ the correctly decodable set

$$\mathcal{B}_n(m_1) = \{y^n \mid y^n = \psi_n(m_1, \phi_{2n}(y^n))\}. \tag{38}$$

We note that since the cardinality of the range of ϕ_{2n} is bounded by $|\mathcal{M}_2|$, then $|\mathcal{B}_n(m_1)| \le |\mathcal{M}_2|$ for all $m_1 \in \mathcal{M}_1$. It can also be seen that

$$P_{\mathrm{WAK}}(\mathrm{error}) = \Pr\{\bar{Y}^n \neq \psi_n(\phi_{1n}(\bar{X}^n), \phi_{2n}(\bar{Y}^n))\} = \Pr\{\bar{Y}^n \notin \mathcal{B}_n(\phi_{1n}(\bar{X}^n))\}. \tag{39}$$

Assume that there are M users in the identification problem where M is specified later. We take ϕ_{1n} as the enrollment mapping for the identification setting. The corresponding enrolled index is denoted by j_i for all $i \in [1\colon M]$. The identification mapping is defined based on the sets $\{\mathcal{B}_n(m_1)\}_{m_1 \in \mathcal{M}_1}$ as follows. We look for a unique index $\hat{w}$ such that $y^n \in \mathcal{B}_n(j_{\hat{w}})$. Otherwise, if there is none or there is more than one such indices, we declare an error. For every $w \in \mathcal{W}$ we then have

$$\begin{aligned}\Pr\{\hat{W} \neq w | W = w\} \le \Pr\{Y^n \notin \mathcal{B}_n(\phi_{1n}(X^n(w))) | W = w\} + \Pr\{Y^n \notin \mathcal{A}_\gamma^n(P_Y) | W = w\} \\ + \Pr\{\exists w' \neq w,\ Y^n \in \mathcal{B}_n(J_{w'}) \cap \mathcal{A}_\gamma^n(P_Y) | W = w\},\end{aligned} \tag{40}$$

where $\mathcal{A}_\gamma^n(P_Y)$ is the weakly typical set based on P_Y and $J_i = \phi_n(X^n(i))$ for all $i \in [1 : M]$. The first term in (40) corresponds to the error expression of the WAK problem and is independent of w. For each $w' \neq w$ we have

$$\begin{aligned}&\Pr\{Y^n \in \mathcal{B}_n(J_{w'}) \cap \mathcal{A}_\gamma^n(P_Y) | W = w\} \\ &= \sum_{j_{w'}} P_{\phi_{1n}(\bar{X}^n)}(j_{w'}) \sum_{y^n \in \mathcal{A}_\gamma^n(P_Y) \cap \mathcal{B}_n(j_{w'})} P_Y^{\otimes n}(y^n) \le e^{-n(H(Y)-\gamma)} |\mathcal{M}_2|.\end{aligned} \tag{41}$$

Therefore we have

$$\Pr\{\hat{W} \neq w | W = w\} \le P_{\mathrm{WAK}}(\mathrm{error}) + \Pr\{\bar{Y}^n \notin \mathcal{A}_\gamma^n(P_Y)\} + M|\mathcal{M}_2| e^{-n(H(Y)-\gamma)}. \tag{42}$$

There exists an $n_0(\gamma) > 0$ such that $\frac{1}{n}\log|\mathcal{M}_2| \le R_2 + \gamma$ for all $n \ge n_0(\gamma)$. If we take $M = e^{n(H(Y)-R_2-3\gamma)}$ then the last term in (42) goes to 0 as $n \to \infty$. Hence,

$$\limsup_{n\to\infty} \Pr\{\hat{W} \neq W\} \le \epsilon. \tag{43}$$

Since γ is arbitrary, this shows that $(R_1, H(Y) - R_2)$ is in $\mathcal{R}_{\mathrm{ID},\epsilon}$.
WAK $\Leftarrow$ ID: Similarly, fix an arbitrary $\gamma > 0$. Given an identification scheme (ϕ_n, ψ_n) such that the rate pair (R_i, R_c) is ϵ-achievable. We construct a code for the WAK-problem as follows. Randomly assign each sequence y^n to a bin $\mathcal{B}(m_2)$ where $m_2 \in [1 : e^{nR_2}]$ and R_2 is specified later. We use the mapping ϕ_n to compress the source sequence x^n at Encoder 1 in the WAK problem. Define the decoding set $\mathcal{D}_n(m_1)$ as

$$\mathcal{D}_n(m_1) = \left\{ y^n \mid \frac{1}{n}\log\frac{P_{\bar{Y}^n|\phi_n(\bar{X}^n)}(y^n|m_1)}{P_{\bar{Y}^n}(y^n)} \ge R_i - 2\gamma,\ y^n \in \mathcal{A}_\gamma^n(P_Y) \right\},\ \forall m_1 \in \mathcal{M}_1. \tag{44}$$

We define a decoder for the WAK-problem as follows. If $\hat{y}^n$ is a unique sequence such that $\hat{y}^n \in \mathcal{B}(m_2) \cap \mathcal{D}_n(m_1)$, then $\hat{y}^n$ is output as the reconstruction sequence. Let $\bar{M}_1$ and $\bar{M}_2$ be the encoded messages at Encoder 1 and 2. Applying Lemma 2 with $M_1 = M$, $P_{X_1^n} = P_X^{\otimes n}$ and $P_{Y^n|X^n} = P_{Y|X}^{\otimes n}$, we obtain

$$\Pr\{\bar{Y}^n \notin \mathcal{D}_n(\bar{M}_1)\} \le \Pr\{\hat{W} \neq W\} + 2e^{-n\gamma} + \Pr\{\bar{Y}^n \notin \mathcal{A}_\gamma^n(P_Y)\}, \tag{45}$$

for all $n \ge n_0(\gamma)$. Furthermore, we have

$$\begin{aligned}&\Pr\{\exists \hat{y}^n \neq \bar{Y}^n, \hat{y}^n \in \mathcal{B}(\bar{M}_2) \cap \mathcal{D}_n(\bar{M}_1)\} \\ &\le \sum_{(m_1, y^n)} P_{\bar{Y}^n \phi_n(\bar{X}^n)}(y^n, m_1) \sum_{m_2} \Pr\{\bar{M}_2 = m_2 | Y^n = y^n\} \sum_{\hat{y}^n \in \mathcal{D}_n(m_1)} \Pr\{\hat{y}^n \in \mathcal{B}(m_2)\} \\ &\le e^{-nR_2} e^{n(H(Y)-R_i+3\gamma)},\end{aligned} \tag{46}$$

[2]The $\mathcal{R}_{\mathrm{WAK},\epsilon}$ also includes all tuples (R_1, R_2) with $R_2 > H(Y)$.

where the last inequality holds due to the following chain of inequalities

$$\begin{aligned}1 &\geq \Pr\{\bar{Y}^n \in \mathcal{D}_n(m_1)|\phi_n(\bar{X}^n) = m_1\}\\ &= \int_{\mathcal{D}_n(m_1)} dP_{\bar{Y}^n|\phi_n(\bar{X}^n)}(y^n|m_1) \geq e^{n(R_i-2\gamma)} \int_{\mathcal{D}_n(m_1)} dP_Y^{\otimes n}(y^n)\\ &\geq e^{n(R_i-3\gamma-H(Y))}|\mathcal{D}_n(m_1)|,\ \forall m_1 \in \mathcal{M}_1,\end{aligned} \tag{47}$$

By setting $R_2 = H(Y) - R_i + 4\gamma$, we have

$$\limsup_{n\to\infty} \Pr\{\hat{\bar{Y}}^n \neq \bar{Y}^n\} \leq \epsilon. \tag{48}$$

Since γ is arbitrary, the rate pair $(R_c, H(Y) - R_i)$ is in $\mathcal{R}_{\mathrm{WAK},\epsilon}$. ■

Remark 5. *In [19] a strong converse proof for the identification problem is provided using Oohama's techniques initiated in [20]. The authors argued in the introduction that classic techniques such as the image characterization [13, Chapter 15] are perhaps not sufficient to establish the strong converse for the identification setting since the identification setting is an unconventional source-channel coding problem. Theorem 6 refutes this point by showing that any technique that can prove the (exponentially) strong converse for the WAK problem automatically proves the (exponentially) strong converse for the identification problem. This includes the classic blowing-up approach in [21] and newly developed approaches in [20], [22] and [23].*

Remark 6. *In Appendix F we present a strong converse proof for $\mathcal{R}_{\mathrm{ID}}$ in both discrete and Gaussian settings. The crucial steps for showing the strong converses in both cases are to use the transformation idea presented in Theorem 6. By noticing that the inequality* (45) *is valid in both discrete and Gaussian cases, we relate the identification error probability to the probability of non-typicality in the decoding of the WAK problem in the discrete setting and the one of a WAK-like problem in the Gaussian setting. In [24] the authors prove the strong converse for a distributed Gaussian source coding with side-information which can be (conceptually) viewed as a different dual of the WAK problem in the Gaussian scenario. After establishing the necessary relation we use the techniques in [24] to show the strong converse for the Gaussian identification problem. Our arguments provide an answer for an open question posed in [19]. In the light of Theorem 6 the proof for the discrete case is not necessary, but we include the arguments therein for completeness.*

Remark 7. *The bound in* (42)*, where the last term is caused by the presence of multiple users in the system, can be extended to both settings in Theorem 1 and Theorem 2 as follows. For simplicity we assume that both settings involve a single group of users, the observation channel is known, and no two input distributions result in the same output distribution. Let $\{(\phi_{1n,s}, \phi_{2n,s}, \psi_{n,s})\}_{s\in\mathcal{S}}$ be an arbitrary set of mappings for the WAK-problems where the corresponding joint distributions are $\{P_{Y|X} \times P_{X,s}\}_{s\in\mathcal{S}}$. A code for the setting of Theorem 2 is constructed as follows. In the enrollment phase we first estimate the state $\hat{s}_i$ and use the corresponding mapping $\phi_{1n,\hat{s}_i}$ to compress the data of the i-th user. In the identification phase the processing unit estimates the underlying state s' and uses the corresponding set $\mathcal{B}_{n,s'}(m_{1,s'}) = \{y^n \mid y^n = \psi_{n,s'}(m_{1,s'}, \phi_{2n,s'}(y^n))\}$ as the decision region. Then we obtain the following upper bound for the setting in Theorem 2*

$$\begin{aligned}&\Pr_{\boldsymbol{s}}\{\hat{W} \neq w|W = w\}\\ &\leq \Pr\{\bar{Y}^n_{s_w} \neq \psi_n(\phi_{1n}(\bar{X}^n_{s_w}), \phi_{2n}(\bar{Y}^n_{s_w}))\} + P(\text{estimation error})\\ &+ \Pr\{\bar{Y}^n_{s_w} \notin \mathcal{A}^n_\gamma(P_{Y_{s_w}})\} + M|\mathcal{M}_{2,s_w}|e^{-n(H(Y_{s_w})-\gamma)}.\end{aligned} \tag{49}$$

The last term in (49) *is valid since given underlying states the independence still holds. From the expression* (49) *we can conclude that two important reasons for the expression* (17) *are vanishing estimation error and the mutual independence of users' data given underlying states.*

Appendix A
Supporting Lemmas

In this section we establish several lemmas that are helpful to prove the results in the paper.

Lemma 1. *Assume that the alphabet $\mathcal{X}$ is a Polish space, specifically, a finite set with discrete metric or $\mathbb{R}$ with Euclidean distance, and $\mathcal{F}$ is the corresponding σ-algebra generated by open sets. Assume that the probability measures $P_{X,s}$ are distinct for all $s \in \mathcal{S}$, i.e., $\forall(s,s') \in \mathcal{S}^2$ such that $s \neq s'$ we have $d(P_{X,s}, P_{X,s'}) \neq 0$ where d is any metric on the space of probability measures. Then there exists a classifier $T : \mathcal{X}^n \to \mathcal{S} \cup \{e\}$, where e denotes an error, such that if $\bar{X}^n \sim P_{X,s}^{\otimes n}$ then*

$$\Pr(T(\bar{X}^n) = s) \to 1,\ \textit{as}\ n \to \infty. \tag{50}$$

Proof: Let d be a metric on the set of probability measures $\mathcal{M}_1(\mathcal{X})$ that induces the weak* topology, for example the Prohorov metric, and define

$$t_X = \inf_{\substack{(s,s')\in\mathcal{S}^2\\ s\neq s'}} d(P_{X,s}, P_{X,s'}). \tag{51}$$

As $|\mathcal{S}|$ is finite, $t_X > 0$. For each x^n the empirical distribution P_{x^n} is given by

$$P_{x^n}(\mathcal{A}) = \frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}(\mathcal{A}), \tag{52}$$

where $\mathcal{A} \in \mathcal{F}$, and δ_x is the corresponding Dirac measure. Define

$$\mathcal{A}_s^X = \{x^n \mid d(P_{x^n}, P_{X,s}) < \frac{t_X}{2}\},\ \forall s \in \mathcal{S}. \tag{53}$$

By the triangle inequality we see that the sets $\mathcal{A}_s^X$ are disjoint. The corresponding classifier is given by

$$T(x^n) \mapsto \begin{cases} s & \text{if } x^n \in \mathcal{A}_s^X \\ e & \text{otherwise.} \end{cases} \tag{54}$$

Furthermore we also see that, if the elements of $\bar{X}^n$ are generated iid from the distribution $P_{X,s}$ then

$$\Pr(\bar{X}^n \notin \mathcal{A}_s^X) \to 0,\ \text{as } n \to \infty \tag{55}$$

due to [25, Theorem 4.4]. ■

Consider an identification setting as in Section II in which users are divided in two groups $[1\colon M_1]$ and $[M_1+1\colon M]$. Data sequences of the first group are identically distributed as $X^n(i) \sim P_{X_1^n}$ for all $i \in [1 : M_1]$. These sequences are mapped into stored indices by a mapping ϕ_{1n}. Similarly data of users in the second group are generated as $X^n(i) \sim P_{X_2^n}$ for all $i \in [M_1+1\colon M]$ and mapped into stored indices by a mapping ϕ_{2n}. The observation channel is given by $P_{Y^n|X^n}$. Then we have the following lemma.

Lemma 2. *Given an arbitrary γ, we have*

$$\begin{aligned}
&\Pr\{\hat{W} \neq W\} \\
&\geq \frac{M_1}{M}\Pr\Big\{\log\frac{dP_{Y_1^n\phi_{1n}(X_1^n)}}{d(P_{Y_1^n}\times P_{\phi_{1n}(X_1^n)})}(Y_1^n, \phi_{1n}(X_1^n)) \leq \log M - n\gamma\Big\} \\
&\quad + \big(1 - \frac{M_1}{M}\big)\Pr\Big\{\log\frac{dP_{Y_2^n\phi_{2n}(X_2^n)}}{d(P_{Y_2^n}\times P_{\phi_{2n}(X_2^n)})}(Y_2^n, \phi_{2n}(X_2^n)) \leq \log M - n\gamma\Big\} - 2e^{-n\gamma},
\end{aligned}$$

where $(Y_1^n, X_1^n) \sim P_{Y^n|X^n}\times P_{X_1^n}$ and $(Y_2^n, X_2^n) \sim P_{Y^n|X^n}\times P_{X_2^n}$.

Proof: For notation simplicity we denote by $\phi(\mathbf{x})$ the tuple $((\phi_{1n}(x^n(i)))_{i=1}^{M_1}, (\phi_{2n}(x^n(i)))_{i=M_1+1}^{M})$ and by Q the product measure $\prod_{i=1}^{M_1} P_{\phi_{1n}(X^n(i))}\times\prod_{i=M_1+1}^{M} P_{\phi_{2n}(X^n(i))}$. Define for each $k \in [1 : M]$ the following correctly decodable and jointly typical sets

$$\begin{aligned}
\mathcal{D}_k &= \{(y^n, \phi(\mathbf{x})) \mid k = \psi_n(y^n, \phi(\mathbf{x}))\} \\
\mathcal{A}_k &= \{(y^n, \phi(\mathbf{x})) \mid \log\frac{dP_{Y_1^n\phi_{1n}(X_1^n)}}{d(P_{Y_1^n}\times P_{\phi_{1n}(X_1^n)})}(y^n, \phi_{1n}(x^n(k))) > \log M - n\gamma\},\ \text{if } k \in [1 : M_1] \\
\mathcal{A}_k &= \{(y^n, \phi(\mathbf{x})) \mid \log\frac{dP_{Y_2^n\phi_{2n}(X_2^n)}}{d(P_{Y_2^n}\times P_{\phi_{2n}(X_2^n)})}(y^n, \phi_{2n}(x^n(k))) > \log M - n\gamma\},\ \text{if } k \in [M_1+1 : M].
\end{aligned} \tag{56}$$

We observe that the sets $\mathcal{D}_k$ are disjoint. We further define the following auxiliary measures

$$\begin{aligned}
Q_k(y^n, \phi(\mathbf{x})) &= P_{Y_1^n\phi_{1n}(X_1^n)}(y^n, \phi_{1n}(x^n(k))\prod_{i=1, i\neq k}^{M_1} P_{\phi_{1n}(X^n(i))}(\phi_{1n}(x^n(i))) \\
&\qquad \times \prod_{i=M_1+1}^{M} P_{\phi_{2n}(X^n(i))}(\phi_{2n}(x^n(i))),\ k \in [1 : M_1], \\
Q_k(y^n, \phi(\mathbf{x})) &= P_{Y_2^n\phi_{2n}(X_2^n)}(y^n, \phi_{2n}(x^n(k))\prod_{i=1}^{M_1} P_{\phi_{1n}(X^n(i))}(\phi_{1n}(x^n(i))) \\
&\qquad \times \prod_{i=M_1+1, i\neq k}^{M} P_{\phi_{2n}(X^n(i))}(\phi_{2n}(x^n(i))),\ k \in [M_1+1 : M].
\end{aligned} \tag{57}$$

Then we have

$$\Pr\{W = \hat{W}\} = \frac{1}{M}\sum_{k=1}^{M} Q_k(\mathcal{D}_k)$$

$$= \frac{1}{M}\sum_{k=1}^{M_1}\Big[\underbrace{Q_k(\mathcal{D}_k\cap\mathcal{A}_k^c)}_{t_{k,11}}+\underbrace{Q_k(\mathcal{D}_k\cap\mathcal{A}_k)}_{t_{k,12}}\Big]+\frac{1}{M}\sum_{k=M_1+1}^{M}\Big[\underbrace{Q_k(\mathcal{D}_k\cap\mathcal{A}_k^c)}_{t_{k,21}}+\underbrace{Q_k(\mathcal{D}_k\cap\mathcal{A}_k)}_{t_{k,22}}\Big]. \tag{58}$$

The terms $t_{k,12}$ and $t_{k,22}$ can be upper bounded using the definition of $\mathcal{A}_k$ by

$$\begin{aligned} t_{k,12} &\le Q_k(\mathcal{A}_k)=\Pr\Big\{\log\frac{dP_{Y_1^n\phi_{1n}(X_1^n)}}{d(P_{Y_1^n}\times P_{\phi_{1n}(X_1^n)})}(Y_1^n,\phi_{1n}(X_1^n))>\log M-n\gamma\Big\}\\ t_{k,22} &\le Q_k(\mathcal{A}_k)=\Pr\Big\{\log\frac{dP_{Y_2^n\phi_{2n}(X_2^n)}}{d(P_{Y_2^n}\times P_{\phi_{2n}(X_2^n)})}(Y_2^n,\phi_{2n}(X_2^n))>\log M-n\gamma\Big\}. \end{aligned} \tag{59}$$

Similarly using change of measure, and definition of $\mathcal{A}_k$, the terms $t_{k,11}$ and $t_{k,21}$ are upper bounded by

$$\begin{aligned} t_{k,11} &\le Me^{-n\gamma}(P_{Y_1^n}\times Q)(\mathcal{D}_k\cap\mathcal{A}_k^c)\le Me^{-n\gamma}(P_{Y_1^n}\times Q)(\mathcal{D}_k)\\ t_{k,21} &\le Me^{-n\gamma}(P_{Y_2^n}\times Q)(\mathcal{D}_k\cap\mathcal{A}_k^c)\le Me^{-n\gamma}(P_{Y_2^n}\times Q)(\mathcal{D}_k). \end{aligned} \tag{60}$$

Therefore we have

$$\begin{aligned} \Pr\{W=\hat{W}\} \le{}& \frac{M_1}{M}\Pr\Big\{\log\frac{dP_{Y_1^n\phi_{1n}(X_1^n)}}{d(P_{Y_1^n}\times P_{\phi_{1n}(X_1^n)})}(Y_1^n,\phi_{1n}(X_1^n))>\log M-n\gamma\Big\}\\ &+\big(1-\frac{M_1}{M}\big)\Pr\Big\{\log\frac{dP_{Y_2^n\phi_{2n}(X_2^n)}}{d(P_{Y_2^n}\times P_{\phi_{2n}(X_2^n)})}(Y_2^n,\phi_{2n}(X_2^n))>\log M-n\gamma\Big\}\\ &+e^{-n\gamma}\underbrace{\Big[(P_{Y_1^n}\times Q)(\bigcup_{k=1}^{M_1}\mathcal{D}_k)+(P_{Y_2^n}\times Q)(\bigcup_{k=M_1+1}^{M}\mathcal{D}_k)\Big]}_{\le 2}. \end{aligned} \tag{61}$$

The conclusion of the lemma follows. ■

Lemma 3. *Assume that $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{T}$ are finite. Define the following function on $\mathbb{R}_+$*

$$\theta(R_c)=\max_{\substack{P_{U|X}:\,|\mathcal{U}|\le|\mathcal{X}|+|\mathcal{T}|\\ I(X;U)\le R_c}}\min_{\tau\in\mathcal{T}} I(Y_\tau;U),\ \textit{where}\ P_{Y_\tau XU}=P_{Y|X,\tau}\times P_{UX},\ \forall\tau\in\mathcal{T}. \tag{62}$$

Then $\theta(R_c)$ is an upper-semicontinuous function, i.e., $\limsup_{R\to R_c}\theta(R)\le\theta(R_c)$.

Proof: Note that since $\mathcal{X}, \mathcal{U}, \mathcal{T}$ and $\mathcal{Y}$ are finite the functions $I(X;U)$ and $I(Y_\tau;U)$ are continuous functions of $P_{U|X}$. Further $P_{U|X}$ can be identified as a vector in $\mathbb{R}^{|\mathcal{U}|\times|\mathcal{X}|}$. Since the set $\{P_{U|X}\mid I(X;U)\le R_c\}$ is closed and bounded, it is a compact subset of $\mathbb{R}^{|\mathcal{U}|\times|\mathcal{X}|}$. The maximization hence can be achieved. The condition for upper semicontinuity states that for an arbitrary $\gamma>0$ there exists an $\epsilon>0$ such that

$$\sup\{\theta(R)\mid |R-R_c|\le\epsilon,\ R\ge 0,\ R\neq R_c\}\le\theta(R_c)+\gamma. \tag{63}$$

Take a decreasing sequence (R_k) such that $R_k\to R_c$ as $k\to\infty$. As we argue previously, for each k there exists a conditional distribution $P_{U|X,k}$ which achieves $\theta(R_k)$. Since $P_{U|X,k}$ all belong to the set $\mathcal{B}_1=\{P_{U|X}\mid I(X;U)\le R_1\}$ which is compact, there exists a subsequence $(P_{U|X,k_l})$ that converges to a conditional distribution $P^\star_{U|X}\in\mathcal{B}_1$. We observe that $\theta(R_c)\le\theta(R_{k_l})=\min_{\tau\in\mathcal{T}}I(Y_\tau;U_{k_l})\to\min_{\tau\in\mathcal{T}}I(Y_\tau;U^\star)$ and $I(X;U^\star)\le R_c$. Therefore we have $\theta(R_c)=\min_{\tau\in\mathcal{T}}I(Y_\tau;U^\star)$, which implies that there exists a $l_0(\gamma)$ such that $\forall l\ge l_0(\gamma)$ we have $\theta(R_{k_l})\le\theta(R_c)+\gamma$. Since $\theta(R_c)$ is a non-decreasing function, setting $\epsilon=R_{k_{l_0}}-R_c$ fulfills the condition (63). ■

Appendix B
Converse of Theorem 1 for finite alphabets

We consider the case that the fraction of users in the first group α is strictly bounded away from 0 and 1, i.e., $\alpha\in(0,1)$. The other cases where $\alpha=0$ or $\alpha=1$ can be derived similarly. Assume that the pair (R_c,R_i) is *achievable*, i.e., there exists an identification scheme such that

$$\begin{aligned} \limsup_{n\to\infty}\frac{1}{n}\log|\mathcal{M}_k|\le R_c,\ k=1,2,\ \liminf_{n\to\infty}\frac{1}{n}\log M\ge R_i,\\ \lim_{n\to\infty}\Pr_{s_1,s_2,\tau}\{\hat{W}\neq W\}=0,\ \forall s_1,s_2,\tau. \end{aligned} \tag{64}$$

Given an arbitrary $\gamma>0$ for all sufficiently large n we have $\log M\ge n(R_i-\gamma)$. Given an arbitrary but unknown triple (s_1,s_2,τ), applying Lemma 2 with $P_{X_1^n}=P_{X,s_1}^{\otimes n}$, $P_{X_2^n}=P_{X,s_2}^{\otimes n}$, and $P_{Y^n|X^n}=P_{Y|X,\tau}^{\otimes n}$, we have

$$\lim_{n\to\infty}\Pr\Big\{\frac{1}{n}\log\frac{dP_{Y_1^n\phi_{1n}(X_1^n)}}{d(P_{Y_1^n}\times P_{\phi_{1n}(X_1^n)})}(Y_1^n,\phi_{1n}(X_1^n))\le R_i-2\gamma\Big\}=0,$$

$$\text{and}\quad \lim_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{dP_{Y_2^n\phi_{2n}(X_2^n)}}{d(P_{Y_2^n}\times P_{\phi_{2n}(X_2^n)})}(Y_2^n,\phi_{1n}(X_2^n)) \le R_i - 2\gamma\right\} = 0. \tag{65}$$

From the definition of *inf-spectral* mutual information, cf. (21), we have

$$R_i \le \underline{I}(\mathbf{Y}_\tau;\phi_k(\mathbf{X}_{s_k})),\ \forall k=1,2,\ \forall(s_k,\tau)\in\mathcal{S}_k\times\mathcal{T} \tag{66}$$

where $(\mathbf{Y}_\tau,\phi_k(\mathbf{X}_{s_k}) = \{(Y_\tau^n,\phi_{kn}(X_{s_k}^n))\}_{n=1}^\infty$ and $(Y_\tau^n,X_{s_k}^n)\sim(P_{Y|X,\tau}\times P_{X,s_k})^{\otimes n}$. For simplicity we use ϕ_n to denote both ϕ_{1n} and ϕ_{2n} in the following. The precise meaning can be inferred from the context. Since the spectral-inf mutual information rate is less than or equal to the inf-mutual information rate [14, Theorem 3.5.2] we have with $\mathcal{S}=\mathcal{S}_1\cup\mathcal{S}_2$

$$R_i \le \underline{I}(\mathbf{Y}_\tau;\phi(\mathbf{X}_s)) \le \liminf_{n\to\infty}\frac{1}{n}I(Y_\tau^n;\phi_n(X_s^n)),\ \forall(s,\tau)\in\mathcal{S}\times\mathcal{T}. \tag{67}$$

Given a positive γ, there exists for each (s,τ) an $n_0(\gamma,s,\tau)$ such that

$$\liminf_{n\to\infty}\frac{1}{n}I(Y_\tau^n;\phi_n(X_s^n)) \le \inf_{n\ge n_0(\gamma,s,\tau)}\frac{1}{n}I(Y_\tau^n;\phi_n(X_s^n))+\gamma. \tag{68}$$

Therefore by considering an $n\ge\max_{s,\tau} n_0(\gamma,s,\tau)$ we have

$$R_i \le \frac{1}{n}I(Y_\tau^n;\phi_n(X_s^n))+\gamma,\ \forall(s,\tau)\in\mathcal{S}\times\mathcal{T}. \tag{69}$$

Now we will apply the standard single-letterization approach. For each (s,τ) we have

$$\begin{aligned} n(R_i-\gamma) \le I(Y_\tau^n;\phi_n(X_s^n)) &= \sum_{i=1}^n I(Y_{\tau,i};Y_\tau^{i-1},\phi_n(X_s^n)) \\ &\le \sum_{i=1}^n I(Y_{\tau,i};X_s^{i-1},\phi_n(X_s^n)). \end{aligned} \tag{70}$$

where the last step holds since $Y_\tau^{i-1} - X_s^{i-1} - (Y_{\tau,i},\phi_n(X_s^n))$ is valid.
Similarly we have

$$n(R_c+\gamma) \ge I(X_s^n;\phi_n(X_s^n)) = \sum_{i=1}^n I(X_{s,i};X_s^{i-1},\phi_n(X_s^n)). \tag{71}$$

Define for each $s\in\mathcal{S}$, $U_{s,i}=(X_s^{i-1},\phi_n(X_s^n))$ for $i\in[1\colon n]$. Let Q be a uniform random variable on $[1:n]$ which is independent of everything. Define further $U_s=(U_{s,Q},Q)$ for each $s\in\mathcal{S}$. Then we obtain, $\forall(s,\tau)\in\mathcal{S}\times\mathcal{T}$

$$\begin{aligned} R_c+\gamma &\ge I(X_{s,Q};U_s) \\ R_i-\gamma &\le I(Y_{\tau,Q};U_s). \end{aligned} \tag{72}$$

Note that

$$P_{Y_{\tau,Q}X_{s,Q}U_s} = P_{Y|X,\tau}\times P_{X,s}\times P_{U_s|X_{s,Q}}. \tag{73}$$

Since each conditional distribution $P_{U|X_{s,Q}}$ acts independently, we can upper bound the cardinality of $\mathcal{U}_s$ by $|\mathcal{X}|+|\mathcal{T}|$ by following [26, Appendix C] as each $P_{U|X_{s,Q}}$ affects $|\mathcal{T}|$ terms $H(Y_{\tau,Q}|U_s)$. This implies that $(R_c+\gamma,R_i-\gamma)\in\bar{\mathcal{R}}_{\text{sc}}$. As $\bar{\mathcal{R}}_{\text{sc}}$ is close, by taking $\gamma\to 0$ we obtain the desired conclusion.

Appendix C
Achievability of Theorem 2

Assume first that $\alpha\in(0,1)$. For each $\kappa\in[1\colon|\mathcal{P}_\mathcal{Y}|]$, we define the following set

$$\mathcal{C}(\kappa) = \{(\tau,s)\mid P_{Y,\kappa} \text{ is the marginal distribution on } \mathcal{Y} \text{ of } P_{Y|X,\tau}\times P_{X,s}\}. \tag{74}$$

For each $s\in\mathcal{S}$ where $\mathcal{S}=\mathcal{S}_1\cup\mathcal{S}_2$ and $\tau\in\mathcal{T}$ assume that the tuple

$$(\bar{X}_s^n,\bar{Y}_\tau^n,\bar{U}_s^n)\sim(P_{UX,s}\times P_{Y|X,\tau})^{\otimes n}.$$

Fix an arbitrary $\delta>0$. Denote $\mathcal{A}_s=\mathcal{A}_\delta^n(P_{XU,s})$, where $\mathcal{A}_\delta^n(P_{XU,s})$ is the weakly typical set corresponding to $P_{XU,s}$. Note that

$$\Pr\{(\bar{X}_s^n,\bar{U}_s^n)\notin\mathcal{A}_s\}\to 0,\ \text{as } n\to\infty. \tag{75}$$

Moreover, define

$$\phi_{s,\tau}(x^n,y^n,u^n) = \mathbf{1}\{(y^n,u^n)\notin\mathcal{A}_\delta^n(P_{Y_\tau U_s})\}, \tag{76}$$

where $P_{Y_\tau U_s}$ is the marginal distribution on $\mathcal{Y} \times \mathcal{U}$ of $P_{Y|X,\tau} \times P_{XU,s}$ then

$$\delta_{n,\tau,s} = \mathbb{E}[\phi_{s,\tau}(\bar{X}_s^n, \bar{Y}_\tau^n, \bar{U}_s^n)] \to 0, \text{ as } n \to \infty, \tag{77}$$

due to the weak law of large numbers. Hence we define

$$\mathcal{B}_{\tau,s} = \{(x^n, u^n) \mid \mathbb{E}[\phi_{\tau,s}(x^n, \bar{Y}_\tau^n, u^n)|\bar{X}_s^n = x^n] \le \delta_{n,\tau,s}^{1/2}\}.$$

Define further $\hat{\mathcal{A}}_s = \mathcal{A}_s \cap \bigcap_\tau \mathcal{B}_{\tau,s}$ and

$$\mathcal{D}(\kappa) = \bigcup_{(\tau,s)\in\mathcal{C}(\kappa)} \mathcal{A}_\delta^n(P_{Y_\tau U_s}). \tag{78}$$

Let T_X be a classifier for s and T_Y be a classifier for κ from Lemma 1.
For each $s \in \mathcal{S}$ generate 2^{nR_c} sequences $u^n(m_s) \sim P_{U_s}$, $m_s \in [1 : e^{nR_c}]$, where P_{U_s} is the marginal corresponding to $P_{UX,s} = P_{U|X,s} \times P_{X,s}$. Hence, we have a total $|\mathcal{S}|e^{nR_c}$ codeword sequences which are used to enroll all users' data.

Enrollment: For each user $i \in [1 : M]$ we first assign $x^n(i)$ to one of the label using the classifier T_X if it is possible. The resulted label is denoted by $\hat{s}_i = T_X(x^n(i))$. Assuming that $\hat{s}_i \neq e$, then we proceed to look for an index $m_{i,\hat{s}_i}$ such that

$$(x^n(i), u^n(m_{i,\hat{s}_i})) \in \hat{\mathcal{A}}_{\hat{s}_i}. \tag{79}$$

Then $m_{i,\hat{s}_i}$ is stored in the database in the i-th position inside the database.

Identification: Given the observation sequence y^n, the processing unit first searches for a suitable label by using the classifier T_Y if it is possible. We denote this label by κ', i.e., $\kappa' = T_Y(y^n)$. If $\kappa' \neq e$, then the processing unit looks for a unique index $\hat{w}$ such that

$$(y^n, u^n(m_{\hat{w},\hat{s}_{\hat{w}}})) \in \mathcal{D}(\kappa') = \bigcup_{(\tau,s)\in\mathcal{C}(\kappa')} \mathcal{A}_\delta^n(P_{Y_\tau U_s}). \tag{80}$$

If there is more than one of such $\hat{w}$ or there is none the system declares an error.
Analysis: For a given but unknown sequence of data states $\boldsymbol{s} = (s_i)_{i=1}^M \in \mathcal{S}_1^{M_1} \times \mathcal{S}_2^{M_2}$ and a channel state τ, we define events $\mathcal{E}_{\text{no}}(w)$, $\mathcal{E}_{\text{es}}(w)$, $\mathcal{E}_{\ge 2}$, $\mathcal{E}_{XU}(w)$, $\mathcal{E}_1(w)$ and $\mathcal{E}_2(w)$ as

$$\begin{aligned}
\mathcal{E}_{\text{es}}(w) &= \{T_X(X^n(w)) \neq s_w\} \cup \{T_Y(Y^n) \neq \kappa_w\}, \text{ where } (s_w, \tau) \in \mathcal{C}(\kappa_w),\\
\mathcal{E}_{XU}(w) &= \{(X^n(w), U^n(m_{w,s_w})) \notin \hat{\mathcal{A}}_{s_w},\ \forall m_{w,s_w}\}\\
\mathcal{E}_1(w) &= \{(Y^n, U^n(M_{w,s_w})) \notin \mathcal{D}(\kappa_w)\},\\
\mathcal{E}_2(w) &= \{\exists w' \neq w \mid (Y^n, U^n(M_{w',\hat{S}_{w'}})) \in \mathcal{D}(\kappa_w)\},\\
\mathcal{E}_{\text{no}}(w) &= \{(Y^n, U^n(M_{w,\hat{S}_w})) \notin \mathcal{D}(T_Y(Y^n))\},\\
\mathcal{E}_{\ge 2}(w) &= \{\exists w' \neq w \mid (Y^n, U^n(M_{w',\hat{S}_{w'}})) \in \mathcal{D}(T_Y(Y^n))\}.
\end{aligned} \tag{81}$$

Assume that the unknown states are $\boldsymbol{s}$, τ and $W = w$, if $\hat{W} \neq w$ then the following event occurs

$$\mathcal{E}_{\text{es}}(w) \cup \mathcal{E}_{\text{no}}(w) \cup \mathcal{E}_{\ge 2}(w).$$

Let $\epsilon > 0$ be arbitrary but given. From Lemma 1 as $X^n(w) \sim P_{X,s_w}^{\otimes n}$, there exists an $n_1(\epsilon, s_w)$ such that if $n \ge n_1(\epsilon, s_w)$ then

$$\Pr\{T_X(X^n(w)) \neq s_w | W = w\} \le \epsilon. \tag{82}$$

Since s_w can be any element of the finite set $\mathcal{S}$, there are only $|S|$ possible values for the left-hand side as w varies, due to the assumption that $X^n(i) \sim P_{X,s_i}^{\otimes n}$. Therefore, if we take $n_1(\epsilon) = \max_{s\in\mathcal{S}} n_1(\epsilon, s)$ then for $n \ge n_1(\epsilon)$

$$\Pr\{T_X(X^n(w)) \neq s_w | W = w\} \le \epsilon,\ \forall s_w \in \mathcal{S},\ \forall w \in [1 : M]. \tag{83}$$

Similarly if $n \ge n_2(\epsilon)$ then

$$\Pr\{T_Y(Y^n) \neq \kappa_w | W = w\} \le \epsilon,\ \forall s_w \in \mathcal{S},\ \forall \tau \in \mathcal{T},\ \forall w \in [1 : M]. \tag{84}$$

Next, we bound the conditional probability of the event $\mathcal{E}_{\text{es}}^c(w) \cap \mathcal{E}_{\text{no}}(w)$ as

$$\begin{aligned}
&\Pr\{\mathcal{E}_{\text{es}}^c(w) \cap \mathcal{E}_{\text{no}}(w) | W = w\}\\
&\le \Pr\{\mathcal{E}_{XU}(w) | W = w\} + \Pr\{\{(X^n(w), U^n(M_{w,s_w})) \in \hat{\mathcal{A}}_{s_w}\} \cap \mathcal{E}_1(w) | W = w\}\\
&= \Pr\{\mathcal{E}_{XU}(w) | W = w\} + \bar{t}_2(w, \boldsymbol{s}, \tau).
\end{aligned} \tag{85}$$

The first term can be bounded as [27, Lemma 5]

$$\Pr\{\mathcal{E}_{XU}(w) | W = w\} \le \Pr\{(\bar{X}_{s_w}^n, \bar{U}_{s_w}^n) \notin \hat{\mathcal{A}}_{s_w}\} + \Pr\left[\log \frac{dP_{\bar{X}_{s_w}^n \bar{U}_{s_w}^n}}{d(P_{\bar{X}_{s_w}^n} \times P_{\bar{U}_{s_w}^n})} \ge n(R_c - \delta/2)\right] + e^{-\exp(n\delta/2)} \le \epsilon, \tag{86}$$

for $n \geq n_3(\epsilon, \delta, s_w)$ if

$$R_c \geq I(X_{s_w}; U_{s_w}) + \delta. \tag{87}$$

Consequently, if we take $n \geq n_3(\epsilon, \delta) = \max_{s \in \mathcal{S}} n_3(\epsilon, \delta, s)$ and $R_c > \max_{s \in \mathcal{S}} I(X_s; U_s) + \delta$ then

$$\Pr\{\mathcal{E}_{XU}(w) | W = w\} \leq \epsilon, \forall s_w \in \mathcal{S}, \ \forall w \in [1 : M]. \tag{88}$$

Furthermore,

$$\begin{aligned} \bar{t}_2(w, \boldsymbol{s}, \tau) &\leq \int_{\hat{\mathcal{A}}_{s_w}} \int_{\mathcal{A}_\delta^n(P_{Y_\tau | U_{s_w}} | u^n)^c} dP_{Y|X,\tau}^{\otimes n}(y^n | x^n) \times dP_{X^n(w) U^n(M_{w,s_w})}(x^n, u^n) \\ &\leq \delta_{n,\tau,s_w}^{1/2} \leq \epsilon, \end{aligned} \tag{89}$$

holds if $n \geq n_4(\epsilon, s_w, \tau)$. Hence taking $n \geq n_4(\epsilon) = \max_{s \in \mathcal{S}, \tau \in \mathcal{T}} n_4(\epsilon, s, \tau)$ we obtain

$$\bar{t}_2(w, \boldsymbol{s}, \tau) \leq \epsilon, \ \forall s_w \in \mathcal{S}, \ \forall \tau \in \mathcal{T}, \ \forall w \in [1 : M]. \tag{90}$$

Next, we have

$$\Pr\{\mathcal{E}_{\text{es}}^c(w) \cap \mathcal{E}_{\geq 2}(w) | W = w\} \leq \Pr\{\mathcal{E}_2(w) | W = w\}. \tag{91}$$

The right-hand side of (91) can be bounded further as follows

$$\begin{aligned} \Pr\{\mathcal{E}_2(w) | W = w\} &\leq \sum_{w' \neq w} \Pr\big\{(Y^n, U^n(M_{w', \hat{S}_{w'}})) \in \mathcal{D}(\kappa_w) | W = w\big\}, \\ &= \sum_{w' \neq w} \sum_{(\tau, s) \in \mathcal{C}(\kappa_w)} \int \int_{\mathcal{A}_\delta^n(P_{Y_\tau | U_s} | u^n)} dP_{Y,\kappa_w}^{\otimes n}(y^n) \times dP_{U^n(M_{w', \hat{S}_{w'}})}(u^n) \\ &\leq \sum_{w' \neq w} \sum_{(\tau, s) \in \mathcal{C}(\kappa_w)} e^{-n(I(Y_\tau; U_s) - 3\delta)} \\ &\leq e^{nR_i} |\mathcal{S}||\mathcal{T}| e^{-n(\min_{s,\tau} I(Y_\tau; U_s) - 3\delta)}. \end{aligned} \tag{92}$$

The second last inequality holds due to the property of weak typicality, which is independent of n. Hence if we take $R_i < \min_{s \in \mathcal{S}, \tau \in \mathcal{T}} I(Y_\tau; U_s) - 3\delta$ and $n \geq n_5(\epsilon, \delta)$ then

$$\Pr\{\mathcal{E}_2(w) | W = w\} < \epsilon, \ \forall s_w \in \mathcal{S}, \ \forall \tau \in \mathcal{T}, \ \forall w \in [1 : M]. \tag{93}$$

In summary, by taking $n > \max\{n_i(\epsilon, \delta)\}_{i=1}^5$ and

$$R_c > \max_{s \in \mathcal{S}} I(X_s; U_s) + \delta, \ \ R_i < \min_{s \in \mathcal{S}, \tau \in \mathcal{T}} I(Y_\tau; U_s) - 3\delta, \tag{94}$$

then

$$\Pr_{\boldsymbol{s},\tau}\{W \neq \hat{W}\} < 6\epsilon, \tag{95}$$

for every tuple $\boldsymbol{s} = (s_i)_{i=1}^M \in \mathcal{S}_1^{M_1} \times \mathcal{S}_2^{M_2}$. Since ϵ and δ are arbitrary, this implies the achievable conclusion for the case $\alpha \in (0, 1)$.

Next we consider the case that $\alpha = 1$. Applying the previous argument directly would give a sub-optimal trade-off region as it gives more extra constraints than needed. To show the achievability in this case, we use random codebook arguments for users in the first group whereas for users in the second group we map their data into a fixed sequence u_0^n. The identification phase works identically as before. Given $\epsilon > 0$ and $\delta > 0$ if we choose $R_c > \max_{s \in \mathcal{S}_1} I(X_s; U_s) + \delta$ and $R_i < \min_{s \in \mathcal{S}_1, \tau \in \mathcal{T}} I(Y_\tau; U_s) - \delta$ then when $n \geq n_0(\delta, \epsilon)$ for some $n_0(\delta, \epsilon)$ sufficiently large we have

$$\Pr_{\boldsymbol{s},\tau}\{\hat{W} \neq w | W = w\} \leq \epsilon, \ \forall w \in [1 : M_1], \quad \forall (\boldsymbol{s}, \tau) \in \mathcal{S}_1^{M_1} \times \mathcal{S}_2^{M_2} \times \mathcal{T}. \tag{96}$$

Therefore the average error probability is upper bounded by

$$\Pr_{\boldsymbol{s},\tau}\{\hat{W} \neq W\} \leq \epsilon + (1 - M_1/M), \ \forall (\boldsymbol{s}, \tau) \in \mathcal{S}_1^M \times \mathcal{S}_2^{M_2} \times \mathcal{T}. \tag{97}$$

Since $\alpha = 1$ and ϵ and δ are arbitrary, we obtain the conclusion in this case. The case $\alpha = 0$ can be solved similarly.

Appendix D
Proof of Theorem 3

Assume that $\alpha \in (0, 1)$.

a) Achievability: For each group $k=1,2$, choose a general source $\mathbf{U}_k$ where U_k^n takes values in a finite alphabet $\mathcal{U}_{k,n}$ such that $(X_k^n, U_k^n) \sim P_{X_k^n U_k^n}$. The condition $\frac{1}{n}\log|\mathcal{U}_{kn}| < N_0,\ k=1,2$, ensures that the right-hand sides in (22) are finite. We hence can assume further that $\mathcal{U}_{1n} = \mathcal{U}_{2n} = \mathcal{U}_n$. For each $k \in \{1,2\}$, let $P_{Y_k^n U_k^n}$ be the marginal distributions of $P_{Y^n|X^n} \times P_{X_k^n U_k^n}$. Define an auxiliary distribution $P_{\bar{Y}^n \bar{U}^n}$ on $\mathcal{Y}_n \times \mathcal{U}_n$ for a general process $(\bar{\mathbf{Y}}, \bar{\mathbf{U}}) = (\bar{Y}^n, \bar{U}^n)_{n=1}^{\infty}$ as

$$P_{\bar{Y}^n \bar{U}^n} = \alpha P_{Y_1^n U_1^n} + (1-\alpha) P_{Y_2^n U_2^n}. \tag{98}$$

Given $\gamma > 0$ we first define the set

$$\mathcal{T}_n = \left\{ (y^n, u^n) \mid \frac{1}{n} \log \frac{P_{\bar{Y}^n|\bar{U}^n}(y^n|u^n)}{P_{\bar{Y}^n}(y^n)} \geq \underline{I}(\bar{\mathbf{Y}}; \bar{\mathbf{U}}) - \gamma \right\}.$$

By [14, Lemma 3.3.1] we have $\underline{I}(\bar{\mathbf{Y}}; \bar{\mathbf{U}}) = \min\{\underline{I}(\mathbf{Y}_1; \mathbf{U}_1), \underline{I}(\mathbf{Y}_2; \mathbf{U}_2)\}$. From the definition of $\mathcal{T}_n$ we have

$$\lim_{n\to\infty} P_{\bar{Y}^n \bar{U}^n}(\mathcal{T}_n^c) = 0, \tag{99}$$

which implies that

$$\lim_{n\to\infty} P_{Y_1^n U_1^n}(\mathcal{T}_n^c) = 0, \text{ and } \lim_{n\to\infty} P_{Y_2^n U_2^n}(\mathcal{T}_n^c) = 0. \tag{100}$$

Then, we define the sets $\mathcal{B}_{kn},\ k=1,2$, as follows

$$\mathcal{B}_{kn} = \left\{ (x^n, u^n) \mid \Pr\{(Y^n, u^n) \notin \mathcal{T}_n | X^n = x^n\} \leq \delta_{kn}^{1/2} \right\},$$

where $\delta_{kn} = \Pr\{(Y_k^n, U_k^n) \notin \mathcal{T}_n\} \to 0$ as $n \to \infty$ for $k=1,2$. By definition of $\mathcal{T}_n$, if $(y^n, u^n) \in \mathcal{T}_n$, then we have

$$\begin{aligned} P_{Y_1^n}(y^n) &\leq \frac{1}{\alpha} e^{-n(\underline{I}(\bar{\mathbf{Y}};\bar{\mathbf{U}})-\gamma)} P_{\bar{Y}^n|\bar{U}^n}(y^n|u^n) \\ P_{Y_2^n}(y^n) &\leq \frac{1}{1-\alpha} e^{-n(\underline{I}(\bar{\mathbf{Y}};\bar{\mathbf{U}})-\gamma)} P_{\bar{Y}^n|\bar{U}^n}(y^n|u^n). \end{aligned} \tag{101}$$

For each group $k,\ k=1,2$, generate a codebook consisting of 2^{nR_c} sequences $u_k^n(m),\ m \in [1:2^{nR_c}]$ where $u_k^n(m) \sim P_{U_k^n}$ and $R_c = \max\{\bar{I}(\mathbf{X}_1; \mathbf{U}_1), \bar{I}(\mathbf{X}_2; \mathbf{U}_2)\} + 2\gamma$. In the enrollment phase for the i-th user which belongs to the k-th group, we look for an index m_i such that $(X^n(i), U_k^n(m_i)) \in \mathcal{B}_{kn}$ and store it in the database. In the identification phase we look for a unique $\hat{w}$ such that $(y^n, u^n(m_{\hat{w}})) \in \mathcal{T}_n$. The rest of the achievability part follows similarly as in the proof of Theorem 2. We highlight some changes in the following. Given that $W = w$ where $w \in [1\colon M_1]$ we have

$$\begin{aligned} &\Pr\{(Y^n, U^n(M_w)) \notin \mathcal{T}_n, (X^n(w), U^n(M_w)) \in \mathcal{B}_{1n} | W = w\} \\ &= \sum_{(x^n,u^n)\in\mathcal{B}_{1n}} \Pr[(Y^n, u^n) \notin \mathcal{T}_n | X^n = x^n] P_{X^n(w)U^n(M_w)}(x^n, u^n) \\ &\leq \sum_{(x^n,u^n)\in\mathcal{B}_{1n}} \delta_{1n}^{1/2} P_{X^n(w)U^n(M_w)}(x^n, u^n) \\ &\leq \delta_{1n}^{1/2}. \end{aligned} \tag{102}$$

Furthermore, we have for $\hat{w} \neq w$,

$$\begin{aligned} \Pr\{(Y^n, U^n(M_{w'})) \in \mathcal{T}_n | W = w\} &= \sum_{(u^n,y^n)\in\mathcal{T}_n} P_{Y_1^n}(y^n) P_{U^n(M_{w'})}(u^n) \\ &\leq \frac{1}{\alpha} e^{-n(\underline{I}(\bar{\mathbf{Y}};\bar{\mathbf{U}})-\gamma)} \sum_{(u^n,y^n)\in\mathcal{T}_n} P_{\bar{Y}^n|\bar{U}^n}(y^n|u^n) P_{U^n(M_{w'})}(u^n) \leq \frac{1}{\alpha} e^{-n(\underline{I}(\bar{\mathbf{Y}};\bar{\mathbf{U}})-\gamma)}. \end{aligned} \tag{103}$$

An analogous analysis can be carried out for $w \in [M_1+1\colon M]$. From the last inequality we can see that $R_i < \underline{I}(\bar{\mathbf{Y}}; \bar{\mathbf{U}}) - \gamma$ suffices to make the error probability to 0 as $n \to \infty$. The cases $\alpha = 0$ and $\alpha = 1$ can be handled similarly as in the proof of Theorem 2.

b) Converse: Assume that the identification-compression rate pair (R_i, R_c) is *achievable*. Then, there exists a triple of identification-compression mappings $(\phi_{1n}, \phi_{2n}, \psi_n)$ such that for every $\gamma > 0$ we have

$$|\mathcal{M}_k| \leq e^{n(R_c+\gamma)},\ k=1,2,\ M \geq e^{n(R_i-\gamma)}, \tag{104}$$

for all $n \geq n_0(\gamma)$. For $k=1,2$, define $U_k^n = \phi_{kn}(X_k^n)$ which takes values on $\mathcal{U}_{kn}$. From (104) it can be seen that for all n, $\frac{1}{n}\log|\mathcal{U}_{kn}| = \frac{1}{n}\log|\mathcal{M}_k|$ is upper bounded by a (large enough) constant. We also have $P_{Y_k^n X_k^n U_k^n} = P_{Y^n|X^n} \times P_{X_k^n U_k^n}$. It can be then shown that

$$R_c + 2\gamma \geq \max\{\bar{I}(\mathbf{X}_k; \mathbf{U}_k)\}_{k=1}^2 \tag{105}$$

along the lines of arguments in [14, p.342]. Using Lemma 2 and taking n to ∞ we observe that

$$R_i \leq \max\{\underline{I}(\mathbf{Y}_k; \mathbf{U}_k)\}_{k=1}^2. \tag{106}$$

Finally taking $\gamma \to 0$ we obtain the claim. The cases $\alpha = 0$ and $\alpha = 1$ can be shown similarly.

Appendix E
Proof of Theorem 5

Let $\mathcal{U}$ be a set such that $|\mathcal{U}| \leq |\mathcal{X}| + |\mathcal{T}|$. For each $s \in \mathcal{S}$, let $P_{U|X,s}$ be the conditional distribution on $\mathcal{U}$ such that $\min_\tau I(Y_\tau; U_s) = \theta^s(R_c)$. Let $\mathbf{U}$ be a discrete process where U^n takes values on the Cartesian product set $\mathcal{U}^n$ such that $P_{X^n U^n} = \sum_{s\in\mathcal{S}} \alpha_s P_{X_s U_s}^{\otimes n}$ where $P_{X_s U_s} = P_{X,s} \times P_{U|X,s}$ for all $s \in \mathcal{S}$. Following the proof of [14, Lemma 3.3.2] we obtain that $\bar{I}(\mathbf{X}; \mathbf{U}) = \sup_{s\in\mathcal{S}} I(X_s; U_s)$. We also have $\underline{I}(\mathbf{Y}; \mathbf{U}) = \inf_{s\in\mathcal{S}} \min_\tau I(Y_\tau; U_s)$.

Applying the above calculation to the result of Theorem 3 with $M_1 = M$ we obtain that

$$\begin{aligned} R_{\max}^{\text{enum}}(R_c) &\geq \inf_{s\in\mathcal{S}} \min_\tau I(Y_\tau; U_s) && (107)\\ &= \inf_{s\in\mathcal{S}} \theta^s(R_c). && (108) \end{aligned}$$

Next we show the other direction. Similarly as in the proof of Theorem 1, given an *achievable* rate pair (R_c, R_i) we obtain for each $s \in \mathcal{S}$ the following inequality for an arbitrary but given $\gamma > 0$ and for all $n \geq n_0(s, \gamma)$

$$\begin{aligned} R_i &\leq \frac{1}{n} I(Y_\tau^n; \phi_{1n}(X_s^n)) + \gamma \leq \frac{1}{n}\sum_{i=1}^n I(Y_{\tau,i}; X_s^{i-1}, \phi_{1n}(X_s^n)) + \gamma \\ &= \frac{1}{n}\sum_{i=1}^n I(Y_{\tau,i}; U_{s,i}) + \gamma = I(Y_{\tau,Q}; U_s) + \gamma,\ \forall \tau \in \mathcal{T}, \end{aligned} \tag{109}$$

where $U_{s,i} = (X_s^{i-1}, \phi_{1n}(X_s^n))$ for all $i \in [1:n]$, Q is an uniform random variable on $[1:n]$ and independent of everything else, and $U_s = (U_{s,Q}, Q)$. We also have $R_c + \gamma \geq I(X_{s,Q}; U_s)$. Therefore we have

$$R_i \leq \sup_{P_{U_s|X_{s,Q}}: R_c+\gamma \geq I(X_{s,Q}; U_s)} \min_{\tau\in\mathcal{T}} I(Y_{\tau,Q}; U_s) + \gamma = \theta^s(R_c + \gamma) + \gamma, \tag{110}$$

since $\mathcal{X}$ and $\mathcal{Y}$ are finite. Furthermore, by Lemma 3, $\theta^s(R_c)$ is an upper-semicontinuous function. Hence for all small enough γ we have $\theta(R_c + \gamma) \leq \theta(R_c) + \gamma_1$ for an arbitrary but given $\gamma_1 > 0$. Therefore $R_i \leq \theta^s(R_c)$ holds for all $s \in \mathcal{S}$. Let $(R_c, R_i) \in \mathcal{R}_{\text{enum}}$ be any rate pair, then there exists a sequence of achievable rate pairs $(R_{c,k}, R_{i,k})$ such that $R_{c,k} \to R_c$ and $R_{i,k} \to R_i$ as $k \to \infty$. Using the given bound we obtain

$$\begin{aligned} R_i &= \limsup_{k\to\infty} R_{i,k} \leq \limsup_{k\to\infty} \inf_{s\in\mathcal{S}} \theta^s(R_{c,k}) \\ &\leq \inf_{s\in\mathcal{S}} \limsup_{k\to\infty} \theta^s(R_{c,k}) \leq \inf_{s\in\mathcal{S}} \theta^s(R_c). \end{aligned} \tag{111}$$

since again $\theta^s(R_c)$ is an upper-semicontinuous function. Therefore $R_{\max}^{\text{enum}}(R_c) \leq \inf_{s\in\mathcal{S}} \theta^s(R_c)$.

Appendix F
A Strong Converse Proof

We present herein strong converse proofs for the discrete and Gaussian settings in the case that both the source distribution P_X and the channel $P_{Y|X}$ are known. In the discrete setting our proof is weaker than in [19] where the authors show the exponential strong converse. We begin with some definitions and important tools.

The definition of ϵ-achievability is already given in (34) in Section V. When $\epsilon = 0$, we denote the corresponding identification-compression trade-off by $\mathcal{R}_{\text{ID}}$. Our proof follows essentially the same arguments in [24, Theorem 4.11] and [21, Theorem 3], where the former uses the following theorem

Theorem 7. *[22, Theorem 9], [24, Corollary 4.7] Consider P_X a probability measure on a finite set $\mathcal{X}$, ν a probability measure on $\mathcal{Y}$ and $P_{Y|X}$. Let $\beta_X = 1/\min_x P_X(x) \in [1, \infty)$, $\alpha = \sup_x \|\frac{dP_{Y|X=x}}{d\nu}\|_\infty \in [1, \infty)$. Let $c \in (0, \infty)$, $\eta, \delta \in (0, 1)$, and $n > 3\beta_X \log \frac{|\mathcal{X}|}{\delta}$. We can choose some set $\mathcal{C}_n$ with $P_X^{\otimes n}[\mathcal{C}_n] \geq 1 - \delta$, such that for $\mu_n = P_X^{\otimes n}\big|_{\mathcal{C}_n}$ we have*

$$\begin{aligned} &\log \mu_n\big(\{x^n : \mathbb{E}_{P_{Y^n|X^n=x^n}}[f] \geq \eta\}\big) - c \log \mathbb{E}_{\nu^{\otimes n}}[f] \\ &\leq n d^\star(P_X, P_{Y|X}, \nu, c) + A\sqrt{n} + c \log \frac{1}{\eta} \end{aligned} \tag{112}$$

for any integrable function f on $\mathcal{Y}^n$ with range in $[0, 1]$, where

$$A = \log(\alpha^c \beta_X^{c+1}) \sqrt{3\beta_X \log \frac{|\mathcal{X}|}{\delta}} + 2c\sqrt{(\alpha - 1)\log \frac{1}{\eta}}. \tag{113}$$

$d^\star$ is defined as[3]

$$d^\star(P_X, P_{Y|X}, \nu, c)$$

[3]The convention is that $\infty - \infty = -\infty$.

$$= \sup_{Q_{UX}:Q_X=P_X} \{cD(Q_{Y|U}||\nu|Q_U) - D(Q_{X|U}||P_X|Q_U)\}.$$

We first examine the discrete case. Let (R_c, R_i) be an ϵ-achievable pair. Then, there exists a pair of mappings (ϕ_n, ψ_n) such that (34) are satisfied. Let $\gamma > 0$ be such that $\epsilon + 3\gamma < 1$. From Lemma 2 with $M_1 = M$, $P_{X_1^n} = P_X^{\otimes n}$ and $P_{Y^n|X^n} = P_{Y|X}^{\otimes}$ we know that for all $n \geq n_0(\gamma)$ we have

$$\begin{aligned}&\epsilon + 2\gamma \\ &\geq \Pr\left\{\frac{1}{n}\log\frac{dP_{Y^n\phi_n(X^n)}}{d(P_{Y^n}\times P_{\phi_n(X^n)})}(Y^n, \phi_n(X^n)) \leq R_i - 2\gamma\right\}.\end{aligned} \tag{114}$$

Then due to the union bound for $n \geq n_1(\gamma)$ we have

$$\Pr\left\{\frac{1}{n}\log\frac{dP_{Y^n\phi_n(X^n)}}{d(P_{Y^n}\times P_{\phi_n(X^n)})}(Y^n, \phi_n(X^n)) \geq R_i - 2\gamma,\ Y^n \in \mathcal{A}_\gamma^n(P_Y)\right\} \geq 1 - \epsilon - 3\gamma. \tag{115}$$

For simplicity we define $\hat{\epsilon} = \epsilon + 3\gamma$. For each $m_1 \in \mathcal{M}_1$ we define the set $\mathcal{D}_n(m_1)$ as in (44). The inequality (115) can be rewritten as

$$\mathbb{E}_{P_X^{\otimes n}}[P_{Y^n|X^n}[\mathcal{D}_n(\phi_n(X^n))]] \geq 1 - \hat{\epsilon}. \tag{116}$$

Choose $\epsilon' \in (\hat{\epsilon}, 1)$ and $\delta = \frac{\epsilon' - \hat{\epsilon}}{2\epsilon'}$. Then by Markov's inequality we have

$$P_X^{\otimes n}\big(\{x^n \mid P_{Y^n|X^n=x^n}[\mathcal{D}_n(\phi_n(x^n))] \geq 1 - \epsilon'\}\big) \geq 1 - \frac{\hat{\epsilon}}{\epsilon'}. \tag{117}$$

We also choose ν as the uniform distribution on $\mathcal{Y}$. Let $c \in (0, \infty)$, $\eta = 1 - \epsilon'$. By Theorem 7 we then can find a measure μ_n such that

$$\mu_n\big(\{x^n \mid P_{Y^n|X^n=x^n}[\mathcal{D}_n(\phi_n(x^n))] \geq \eta\}\big) \geq 1 - \frac{\hat{\epsilon}}{\epsilon'} - \delta = \delta, \tag{118}$$

and

$$\begin{aligned}&\log\mu_n\big(\{x^n \mid \mathbb{E}_{P_{Y^n|X^n=x^n}}[f] \geq \eta\}\big) - c\log\mathbb{E}_{\nu^{\otimes n}}[f] \\ &\quad\leq nd^\star(P_X, P_{Y|X}, \nu, c) + \mathcal{O}\big(\sqrt{n}\big),\end{aligned} \tag{119}$$

for any integrable f with range in $[0, 1]$. Further calculation indicates that

$$\begin{aligned}&d^\star(P_X, P_{Y|X}, \nu, c) \\ &= \sup_{\substack{U:U-X-Y \\ I(X;U)<\infty}} \{-cH(Y|U) - I(X;U)\} + c\log|\mathcal{Y}| \\ &= \max_{\substack{U:U-X-Y \\ |\mathcal{U}|\leq|\mathcal{X}|+1}} \{-cH(Y|U) - I(X;U)\} + c\log|\mathcal{Y}|\end{aligned} \tag{120}$$

due to the Support Lemma [13, Lemma 15.4]. Since there are $|\mathcal{M}_1|$ possible values of m_1, it follows that there must exists $m_1^\star$ such that

$$\mu_n\big(\{x^n \mid P_{Y^n|X^n=x^n}[\mathcal{D}_n(m_1^\star)] \geq \eta\}\big) \geq \frac{\delta}{|\mathcal{M}_1|}. \tag{121}$$

Moreover we also have from (47) that

$$\nu^{\otimes n}(\mathcal{D}_n(m_1^\star)) \leq |\mathcal{Y}|^{-n} e^{n(H(Y)+3\gamma-R_i)}. \tag{122}$$

Combining (120), (121) and (122), and taking f to be the indicator function of the set $\mathcal{D}_n(m_1^\star)$ we obtain

$$\begin{aligned}&-\log|\mathcal{M}_1| - cn(-\log|\mathcal{Y}| + H(Y) + 3\gamma - R_i) \\ &\leq n\left(\max_{\substack{U:U-X-Y \\ |\mathcal{U}|\leq|\mathcal{X}|+1}} \{-cH(Y|U) - I(X;U)\} + c\log|\mathcal{Y}|\right) + \mathcal{O}(\sqrt{n}).\end{aligned} \tag{123}$$

This implies further that

$$\begin{aligned}&R_c - cR_i + (1+3c)\gamma \\ &\geq \min_{\substack{U:U-X-Y \\ |\mathcal{U}|\leq|\mathcal{X}|+1}} \{I(X;U) - cI(Y;U)\} + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right) \\ &= f(c) + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right).\end{aligned} \tag{124}$$

Taking $n \to \infty$ we see that

$$(R_c+\gamma, R_i-3\gamma) \in \bigcap_{c>0}\{R_a - cR_b \geq f(c)\} \overset{\text{(hp)}}{=} \mathcal{R}_{\text{ID}}, \tag{125}$$

where (hp) follows from the hyper plane characterization of a closed convex set which is explained in details at the end of this section. Hence, this leads to $\mathcal{R}_{\text{ID},\epsilon} \subseteq \mathcal{R}_{\text{ID}}$.

In the Gaussian case, assume that $P_X = \mathcal{N}(0, \sigma_X^2)$ and $P_{Y|X=x} = \mathcal{N}(x, 1)$, then by doing the same steps, with ν in this case the Lebesgue measure, we obtain (119) using [24, Corollary 4.10]. Since ν is the Lebesgue measure, similarly as in (47) we have

$$\nu(\mathcal{D}_n(m^\star)) \leq e^{n(h(Y)+3\gamma-R_i)}. \tag{126}$$

Additionally,

$$\begin{aligned} d^\star(P_X, P_{Y|X}, \nu, c) &= \sup_{\substack{U:U-X-Y\\ I(X;U)<\infty}} \{-c \times h(Y|U) - I(X;U)\} \\ &\overset{(**)}{\leq} \sup_{1<\beta\leq 1+\sigma_X^2} \Big\{ -\frac{c}{2}\log 2\pi e\beta - \frac{1}{2}\log\frac{\sigma_X^2}{\beta-1} \Big\}, \end{aligned} \tag{127}$$

where $(**)$ follows by first putting $h(Y|U) = \frac{1}{2}\log 2\pi e \beta$ and then using the entropy power inequality. Therefore we obtain,

$$\begin{aligned} & R_c - cR_i + (1+3c)\gamma \\ &\geq \inf_{0\leq\beta<\log(1+\sigma_X^2)} \frac{1}{2}\Big\{ \log\frac{\sigma_X^2}{(\sigma_X^2+1)e^{-\beta}-1} - c\beta \Big\} \\ &= \begin{cases} 0 & \text{if } 0\leq c\leq 1 \text{ at } \beta = 0 \\ \frac{1}{2}\big\{ \log\sigma_X^2(c-1) & \text{if } c>1 \\ \quad -c\log(\sigma_X^2+1)(1-1/c)\big\} & \text{at } \beta = \log(\sigma_X^2+1)(1-1/c) \end{cases}. \end{aligned} \tag{128}$$

Compared with the characterization in (13), we observe that

$$(R_c+\gamma, R_i-3\gamma) \overset{\text{(hp)}}{\in} \mathcal{R}_{\text{ID}}. \tag{129}$$

Therefore, we have $\mathcal{R}_{\text{ID},\epsilon} \subseteq \mathcal{R}_{\text{ID}}$.

On (hp): The inclusion $\bigcap_{c>0}\{R_a - cR_b \geq f(c)\} \supseteq \mathcal{R}_{\text{ID}}$ is quite straightforward. Since $\mathcal{R}_{\text{ID}}$ is a closed convex subset of $\mathbb{R}_+^2$ if $(x,y) \in \mathbb{R}_+^2$ and $(x,y) \notin \mathcal{R}_{\text{ID}}$ then there exists a vector $(a,b) \in \mathbb{R}^2$ such that

$$ax + by < aR_1 + bR_2,\ \forall (R_1, R_2) \in \mathcal{R}_{\text{ID}}. \tag{130}$$

Since $(0,0) \in \mathcal{R}_{\text{ID}}$, we see that either a or b must be negative. If $a<0$, then plugging $(R_1, 0)$ in the above inequality we obtain the violation for sufficiently large R_1. Hence, we have $a>0$ and $b<0$. We can normalize further to obtain

$$x - cy < R_1 - cR_2,\ \forall (R_1, R_2) \in \mathcal{R}_{\text{ID}}, \tag{131}$$

where $c = -b/a > 0$. The minimum of the right-hand side, which is attained by a point $(R_1^*, R_2^*) \in \mathcal{R}_{\text{ID}}$, is $f(c)$. Therefore if $(x,y) \notin \mathcal{R}_{\text{ID}}$ then $(x,y) \notin \bigcap_{c>0}\{R_a - cR_b \geq f(c)\}$, which implies that $\bigcap_{c>0}\{R_a - cR_b \geq f(c)\} \subseteq \mathcal{R}_{\text{ID}}$.

References

[1] F. Willems, T. Kalker, and J.-P. Linnartz, "On the capacity of a biometrical identification system," in *2003 IEEE International Symposium on Information Theory*. IEEE, p. 82.

[2] F. Farhadzadeh, S. Voloshynovskiy, and O. Koval, "Performance analysis of content-based identification using constrained list-based decoding," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 5, pp. 1652–1667, 2012.

[3] T. Ignatenko and F. M. Willems, "Fundamental limits for privacy-preserving biometric identification systems that support authentication," *IEEE Trans. Inf. Theory*, vol. 61, no. 10, pp. 5583–5594, 2015.

[4] E. Tuncel, "Capacity/storage tradeoff in high-dimensional identification systems," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2097–2106, 2009.

[5] M. B. Westover and J. A. O'Sullivan, "Achievable rates for pattern recognition," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 299–320, 2008.

[6] E. Tuncel and D. Gündüz, "Identification and lossy reconstruction in noisy databases," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 822–831, 2014.

[7] F. M. J. Willems, "Searching methods for biometric identification systems: Fundamental limits," in *2009 IEEE International Symposium on Information Theory*. IEEE, 2009, pp. 2241–2245.

[8] E. Tuncel, "Recognition capacity versus search speed in noisy databases," in *2012 IEEE International Symposium on Information Theory*. IEEE, pp. 2566–2570.

[9] F. Farhadzadeh and F. M. Willems, "Identification rate, search and memory complexity tradeoff: Fundamental limits," *IEEE Trans. Inf. Theory*, vol. 62, no. 11, pp. 6173–6188, 2016.

[10] D. Blackwell, L. Breiman, and A. Thomasian, "The capacity of a class of channels," *The Annals of Mathematical Statistics*, pp. 1229–1241, 1959.

[11] ——, "The capacities of certain channel classes under random coding," *The Annals of Mathematical Statistics*, vol. 31, no. 3, pp. 558–567, 1960.

[12] R. Ahlswede, “Elimination of correlation in random codes for arbitrarily varying channels,” *Probability Theory and Related Fields*, vol. 44, no. 2, pp. 159–175, 1978.
[13] I. Csiszar and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.
[14] T. S. Han, *Information-Spectrum Methods in Information Theory*. Springer-Verlag Berlin Heidelberg, 2003.
[15] R. Ahlswede and J. Körner, “Source coding with side information and a converse for degraded broadcast channels,” *IEEE Trans. Inf. Theory*, vol. 21, no. 6, pp. 629–637, 1975.
[16] A. Wyner, “On source coding with side information at the decoder,” *IEEE Trans. Inf. Theory*, vol. 21, no. 3, pp. 294–300, 1975.
[17] V. Yachongka and H. Yagi, “Reliability function and strong converse of biomedical identification systems,” in *2016 International Symposium on Information Theory and Its Applications (ISITA)*. IEEE, 2016, pp. 547–551.
[18] C. Tian and J. Chen, “Successive refinement for hypothesis testing and lossless one-helper problem,” *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4666–4681, 2008.
[19] L. Zhou, V. Y. Tan, L. Yu, and M. Motani, “Exponential strong converse for content identification with lossy recovery,” *IEEE Trans. Inf. Theory*, vol. 64, no. 8, pp. 5879–5897, 2018.
[20] Y. Oohama, “Exponential strong converse for one helper source coding problem,” *Entropy*, vol. 21, no. 6, p. 567, 2019.
[21] R. Ahlswede, P. Gács, and J. Körner, “Bounds on conditional probabilities with applications in multi-user communication,” *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 34, no. 2, pp. 157–177, 1976.
[22] J. Liu, R. van Handel, and S. Verdú, “Beyond the blowing-up lemma: Sharp converses via reverse hypercontractivity,” in *2017 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2017, pp. 943–947.
[23] S. Watanabe, “A converse bound on wyner-ahlswede-körner network via gray-wyner network,” in *2017 IEEE Information Theory Workshop (ITW)*. IEEE, 2017, pp. 81–85.
[24] J. Liu, R. van Handel, and S. Verdú, “Second-order converses via reverse hypercontractivity,” *arXiv preprint arXiv:1812.10129*, 2018.
[25] P. Mitran, “On a Markov Lemma and Typical Sequences for Polish Alphabets,” *IEEE Trans. Inf. Theory*, vol. 61, no. 10, pp. 5342–5356, 2015.
[26] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge university press, 2011.
[27] S. Verdú, “Non-asymptotic achievability bounds in multiuser information theory,” in *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*. IEEE, 2012, pp. 1–8.