Mathematics and Engineering Communications Laboratory

Technical Report

Inner and Outer Bounds for the Public Information Embedding Capacity Region Under Multiple Access Attacks

Y. Zhong, *Y. Wang*, *F. Alajaji*, and *T. Linder*

March 2010

Inner and Outer Bounds for the Public Information Embedding Capacity Region Under Multiple Access Attacks*

Yangfan Zhong Yadong Wang Fady Alajaji Tamás Linder

March 5, 2010

Abstract

We consider a public multi-user information embedding (watermarking) system in which two messages (watermarks) are independently embedded into two correlated covertexts and are transmitted through a multiple-access attack channel. The tradeoff between the achievable embedding rates and the average distortions for the two embedders is studied. For given distortion levels, inner and outer bounds for the embedding capacity region are obtained in single-letter form. Tighter bounds are also given for independent covertexts.

Index Terms: Capacity region, correlated covertexts, multiple access attack, multi-user information embedding, inner and outer bounds, public watermarking.

1 Introduction

In the last decade, the single-user (point-to-point) information-hiding (information-embedding, watermarking) model has been thoroughly studied from an information-theoretic point of view; see, e.g., [1, 9, 15] and the references therein. With the rapid development of wired and wireless communication networks, situations arise where privacy protection is no longer a point-to-point problem. Therefore, it is of interest to study information-hiding problems in multi-user settings.

In this paper we consider the scenario in which two secret messages (watermarks) are independently embedded in two correlated sources (covertexts) and are then jointly decoded under multiple-access attacks. This scenario is motivated by, for example, the practical situation where audio and video frames are watermarked separately, but they are transmitted in a single bit stream and decoded by one multimedia player (see [10, 12, 8]). The model is depicted in Fig. 1 and it assumes that two users separately embed their watermarks W_1 and W_2 into two correlated discrete memoryless sources (DMSs), U_1 and U_2. Each user can only access one of the two covertexts. The watermarked messages (stegotexts) X_1^n and X_2^n are then

*This research was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC). The material in this correspondence was presented in part at the IEEE International Symposium on Information Theory, Toronto, July 2008.

Yangfan Zhong and Yadong Wang are with the Bank of Montreal, 8th floor, 302 Bay St., Toronto, Canada (email: zhongyangfan@hotmail.com, y.d.wang99@gmail.com); Fady Alajaji and Tamás Linder are with the Department of Mathematics & Statistics, Queen's University, Kingston, ON K7L 3N6, Canada (email: {fady,linder}@mast.queensu.ca).

sent through a multiple-access attack channel (MAAC) to a decoder which attempts to reconstruct the watermarks. For this two-user information embedding system we are interested in determining the embedding capacity region; i.e., the two-dimensional set of all achievable embedding rate pairs under constraints on the embedding distortions.

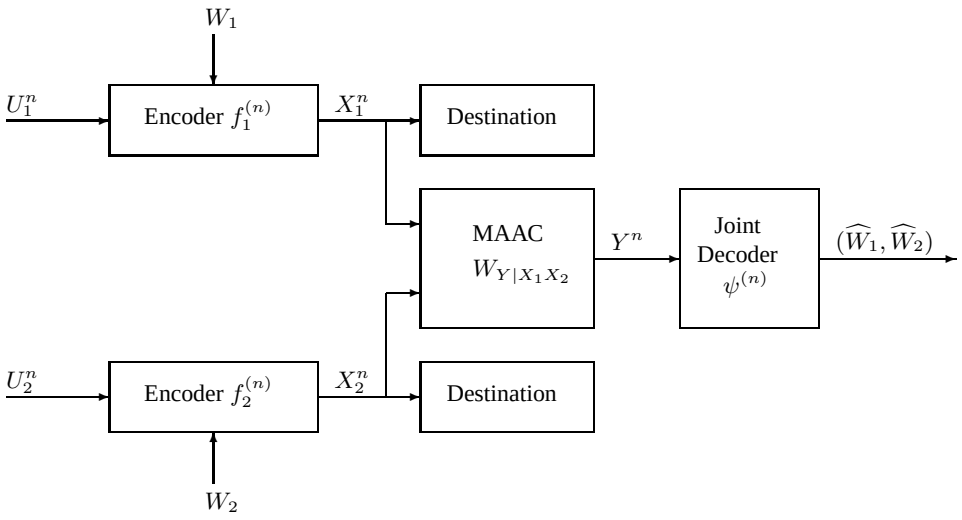

Figure 1: A multi-user information embedding system with two embedders.

Our main result (Theorem 1) is an inner bound for the embedding capacity region. The proof is based on the approach of Gelfand and Pinsker [5] and a strong typicality coding/decoding argument. The encoders first map the watermarks W_1 and W_2 and the correlated covertexts U_1^n and U_2^n to auxiliary codewords T_1^n and T_2^n, and then generate two stegotexts X_1^n and X_2^n which are jointly typical with $(U_1^n, U_2^n, T_1^n, T_2^n)$. The decoder recovers the watermarks by examining the joint typicality of the received sequence Y^n and all auxiliary codeword pairs (T_1^n, T_2^n).

One major technical difficulty is the problem of how to separately construct the typical sequence encoders. In order to guarantee that the codewords together with the covertexts are jointly typical with a high probability, we adopt a "Markov" encoding scheme from [11], which was originally proposed for Gaussian multi-terminal source coding (see also [13] and [6]). The Markov encoders can be briefly described as follows. One of the encoders (embedders), say Encoder 1, first forms an estimate of the source sequence of the other encoder, and then generates T_1^n which is jointly typical with the observed source sequence U_1^n and the estimated source sequence. The other encoder, Encoder 2, first forms an estimate of the source sequence as well as the auxiliary codeword of Encoder 1, and then generates T_2^n which is jointly typical with the source sequence U_2^n and all the other sequences estimated. For the resulting scheme, an extended Markov lemma (Lemma 3) ensures that the auxiliary codewords T_1^n and T_2^n, although generated by separate encoders, are jointly typical with the source sequences with a high probability.

We also derive an outer bound for the embedding capacity region with single-letter characterization (Theorem 2), using Fano's inequality and a standard information-theoretical bounding argument. We specialize the embedding capacity region to independent covertexts and obtain inner and outer bounds for this case (Theorem 3). The inner bound is a consequence of Theorem 1, while in the converse part we sharpen the bound of Theorem 2 by making use of the independence condition.

We note that the multi-user information embedding problem studied in this paper is related to the works [12] and [8]. In [12], the authors present an achievable embedding region for correlated Gaussian covertexts and parallel (independent) additive Gaussian attack channels (as opposed to the MAAC considered here). In a recent work [8], the authors study the same system as ours and give an inner bound for the capacity region without a proof, stating that this inner bound can be easily proved via the coding procedure in [12]. However, the proof in [12] seems to be incorrect because the encoders cannot guarantee the typicality of the output sequences with respect to the covertexts sequences. Our code construction corrects this problem and in Theorem 1 we show that the main result in [12] (the achievable region) and the inner bound given in [8] are both correct. We also point out that a similar setup concerning a multi-user reversible information embedding system was considered in [7] and [8] for two covertexts and a MAAC. Since in the reversible information embedding problem the secret messages and the covertexts are both reconstructed at the decoder, Gelfand and Pinsker coding is not required and the coding strategy is fundamentally different from ours.

The remainder of this paper is organized as follows. We set up the public multi-user embedding (watermarking) problem, define the embedding capacity region, and present our main results in Section 2. The proof of the inner bound is given in Section 3, while the proof of the outer bounds are deferred to the Appendix. We close the paper with concluding remarks in Section 4.

2 Problem Formulation and Main Results

Let $|\mathcal{X}|$ denote the size of a finite set $\mathcal{X}$. If X is a random variable (RV) with distribution P_X, we denote its n-dimensional product distribution by $P_X^{(n)}$. Similar notation applies to joint and conditional distributions. For RVs X, Y, and Z with joint distribution P_{XYZ}, we use P_X, P_{XY}, $P_{YZ|X}$, etc., to denote the corresponding marginal and conditional probabilities induced by P_{XYZ}. The expectation of the RV X is denoted by $\mathbb{E}(X)$. All alphabets are finite, and all logarithms and exponentials are in base 2.

Let U_1 and U_2 be two discrete memoryless host sources with alphabets $\mathcal{U}_1$ and $\mathcal{U}_2$ and joint distribution $Q_{U_1U_2}$. The watermarks W_1 and W_2 are independently and uniformly chosen from the sets $\mathcal{W}_1 \triangleq \{1, 2, ..., M_1\}$ and $\mathcal{W}_2 \triangleq \{1, 2, ..., M_2\}$, respectively. The attack channel is modeled as a two-sender one-receiver discrete memoryless MAAC $W_{Y|X_1X_2}$ having input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$, output alphabet $\mathcal{Y}$, and transition probability distribution $W_{Y|X_1X_2}(y|x_1, x_2)$. The probability of receiving $\mathbf{y} \in \mathcal{Y}^n$ conditioned on sending $\mathbf{x}_1 \in \mathcal{X}_1^n$ and $\mathbf{x}_2 \in \mathcal{X}_2^n$ is hence given by $W_{Y|X_1X_2}^{(n)}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)$.

Let $d_i : \mathcal{U}_i \times \mathcal{X}_i \to [0, \infty)$ be single-letter distortion measures and define $d_i^{max} \triangleq \max_{u_i, x_i} d_i(u_i, x_i)$ for $i = 1, 2$. For $\mathbf{u}_i \in \mathcal{U}_i^n$ and $\mathbf{x}_i \in \mathcal{X}_i^n$, let $d_i(\mathbf{u}_i, \mathbf{x}_i) = \sum_{j=1}^n d_i(u_{ij}, x_{ij})$.

A two-sender one-receiver multiple-access embedding (MAE) code $(f_1^{(n)}, f_2^{(n)}, \psi^{(n)})$ with block length n consists of (see Fig. 1) two encoders (embedders)

$$f_1^{(n)} : \mathcal{W}_1 \times \mathcal{U}_1^n \longrightarrow \mathcal{X}_1^n \quad \text{and} \quad f_2^{(n)} : \mathcal{W}_2 \times \mathcal{U}_2^n \longrightarrow \mathcal{X}_2^n$$

with embedding rates $R_{f_1} = \frac{1}{n} \log_2 M_1$ and $R_{f_2} = \frac{1}{n} \log_2 M_2$, respectively, and a decoder

$$\psi^{(n)} : \mathcal{Y}^n \longrightarrow \mathcal{W}_1 \times \mathcal{W}_2.$$

The system depicts a "public" embedding scenario since the covertexts are not available at the decoder.

The probability of erroneously decoding the secret messages is given by

$$\begin{aligned} P_e^{(n)} &\triangleq \Pr\big(\psi^{(n)}(Y^n) \neq (W_1, W_2)\big) \\ &= \frac{1}{2^{n(R_1+R_2)}} \sum_{w_1=1}^{M_1} \sum_{w_2=1}^{M_2} \sum_{\mathcal{U}_1^n \times \mathcal{U}_2^n} Q_{U_1U_2}^{(n)}(\mathbf{u}_1, \mathbf{u}_2) W_{Y|X_1X_2}^{(n)} \Big(\mathbf{y} : \psi^{(n)}(\mathbf{y}) \neq (w_1, w_2) | \mathbf{x}_1, \mathbf{x}_2\Big) \end{aligned}$$

where $\mathbf{x}_i \triangleq f_i^{(n)}(w_i, \mathbf{u}_i)$ for $i = 1, 2$.

Definition 1 Given $Q_{U_1U_2}$, $W_{Y|X_1X_2}$, a rate pair (R_1, R_2) is said to be achievable with respect to distortion levels (D_1, D_2) if there exists a sequence of MAE codes $(f_1^{(n)}, f_2^{(n)}, \psi^{(n)})$ at embedding rates no smaller than R_1 and R_2, respectively, such that $\lim_{n\to\infty} P_e^{(n)} = 0$ and

$$\limsup_{n\to\infty} \frac{1}{n} \mathbb{E}\left[d_i(U_i^n, f_i^{(n)}(W_i, U_i^n))\right] \leq D_i, \quad i = 1, 2.$$

The embedding capacity region $\mathcal{R}(D_1, D_2)$ is the closure of the set of all achievable rate pairs (R_1, R_2).

Remark 1 It can be shown by using a time-sharing argument [4] that $\mathcal{R}(D_1, D_2)$ is convex.

Definition 2 Given $Q_{U_1U_2}$, $W_{Y|X_1X_2}$, and a pair of distortion levels (D_1, D_2), let $\mathcal{S}_{D_1,D_2}$ be the set of RVs $(U_1, T_1, U_2, T_2, X_1, X_2, Y) \in \mathcal{U}_1 \times \mathcal{T}_1 \times \mathcal{U}_2 \times \mathcal{T}_2 \times \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}$ for some finite alphabets $\mathcal{T}_1$ and $\mathcal{T}_2$ such that the joint distribution $P_{U_1T_1U_2T_2X_1X_2Y}$ satisfies: (1) $P_{U_1T_1U_2T_2X_1X_2Y} = Q_{U_1U_2} P_{T_1X_1|U_1} P_{T_2X_2|U_2} W_{Y|X_1X_2}$, (2) $I(U_i; T_i) > 0$, and (3) $\mathbb{E}[d_i(U_i, X_i)] \leq D_i$, for $i = 1, 2$.

Definition 3 Given $Q_{U_1U_2}$, $W_{Y|X_1X_2}$, and a pair of distortion levels (D_1, D_2), let $\mathcal{P}_{D_1,D_2}$ be the set of RVs $(U_1, T_1, U_2, T_2, X_1, X_2, Y) \in \mathcal{U}_1 \times \mathcal{T}_1 \times \mathcal{U}_2 \times \mathcal{T}_2 \times \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}$ for some finite alphabets $\mathcal{T}_1$ and $\mathcal{T}_2$ such that the joint distribution $P_{U_1T_1U_2T_2X_1X_2Y}$ satisfies: (1) $P_{U_1T_1U_2T_2X_1X_2Y} = Q_{U_1U_2} P_{T_1T_2X_1X_2|U_1U_2} W_{Y|X_1X_2}$, and (2) $\mathbb{E}[d_i(U_i, X_i)] \leq D_i$, for $i = 1, 2$.

Note that the only difference between the two regions is that in the definition of $\mathcal{S}_{D_1,D_2}$, the conditional distribution of (T_1, T_2, X_1, X_2) given (U_1, U_2) is restricted to be in the form $P_{T_1X_1|U_1} P_{T_2X_2|U_2}$. This of course implies $\mathcal{S}_{D_1,D_2} \subseteq \mathcal{P}_{D_1,D_2}$.

The following are the main results of the paper.

Theorem 1 (Inner bound) Let $\mathcal{R}_{in}(D_1, D_2)$ be the closure of the convex hull of all (R_1, R_2) satisfying

$$\begin{aligned} R_1 &< I(T_1; T_2, Y) - I(U_1; T_1), && (1) \\ R_2 &< I(T_2; T_1, Y) - I(U_2; T_2), && (2) \\ R_1 + R_2 &< I(T_1, T_2; Y) - I(U_1, U_2; T_1, T_2), && (3) \end{aligned}$$

for some $(U_1, T_1, U_2, T_2, X_1, X_2, Y) \in \mathcal{S}_{D_1,D_2}$. Then $\mathcal{R}_{in}(D_1, D_2) \subseteq \mathcal{R}(D_1, D_2)$.

The proof of the theorem is given in Section 3.

Remark 2 As we show in Appendix C, the cardinality of the alphabets of the auxiliary RVs T_1 and T_2 for $\mathcal{R}_{in}(D_1, D_2)$ can be bounded as $|\mathcal{T}_i| \leq |\mathcal{U}_1||\mathcal{U}_2||\mathcal{X}_i| + 1$, $i = 1, 2$.

Remark 3 Although we only deal with discrete (finite-alphabet) sources and channels, it is not hard to see that, with the appropriate changes in the proof, the achievable region is also valid for a system that incorporates a pair of correlated memoryless Gaussian sources and a Gaussian MAAC. In particular, when the MAAC is a pair of parallel (independent) additive Gaussian channels, $\mathcal{R}_{in}(D_1, D_2)$ is the achievable region obtained in [12], even though the proof provided in [12] is not entirely correct. Note also that our inner bound $\mathcal{R}_{in}(D_1, D_2)$ is the same as the one given without proof in [8, Proposition 1].

Theorem 2 (Outer bound) Let $\mathcal{R}_{out}(D_1, D_2)$ be the closure of the collection of all rate pairs (R_1, R_2) satisfying conditions (1)–(3) for some $(U_1, T_1, U_2, T_2, X_1, X_2, Y) \in \mathcal{P}_{D_1,D_2}$. Then $\mathcal{R}(D_1, D_2) \subseteq \mathcal{R}_{out}(D_1 + \delta, D_2 + \delta)$ for all $\delta > 0$.

The proof of the theorem is given in Appendix A. The proof involves Fano's inequality and a (by now) rather standard information-theoretic argument that generalizes the converse proof for a single-user embedding system in [15].

Remark 4 The above theorem states that $\mathcal{R}(D_1, D_2) \subseteq \bigcap_{\delta>0} \mathcal{R}_{out}(D_1 + \delta, D_2 + \delta)$. If we could upper bound the cardinality of the alphabet sizes of the auxiliary RVs T_1 and T_2 in the definition of $\mathcal{R}_{out}(D_1, D_2)$, it would be easy to show that $\bigcap_{\delta>0} \mathcal{R}_{out}(D_1 + \delta, D_2 + \delta) = \mathcal{R}_{out}(D_1, D_2)$, so that $\mathcal{R}(D_1, D_2) \subseteq \mathcal{R}_{out}(D_1, D_2)$. However, without such an upper bound, we can only state the theorem in the present weaker form. The same remark applies to the outer bound in the next theorem.

We next consider the special case when the covertexts are independent; i.e., $Q_{U_1U_2} = Q_{U_1}Q_{U_2}$. We then have the following inner and outer bounds.

Theorem 3 Let $Q_{U_1U_2} = Q_{U_1}Q_{U_2}$. Let $\mathcal{R}^*_{in}(D_1, D_2)$ be the closure of the convex hull of all (R_1, R_2) satisfying

$$R_1 < I(T_1; Y|T_2) - I(U_1; T_1) \quad (4)$$

$$R_2 < I(T_2; Y|T_1) - I(U_2; T_2) \quad (5)$$

$$R_1 + R_2 < I(T_1, T_2; Y) - I(U_1; T_1) - I(U_2; T_2) \quad (6)$$

for some $(U_1, T_1, U_2, T_2, X_1, X_2, Y) \in \mathcal{S}_{D_1,D_2}$, and let $\mathcal{R}^*_{out}(D_1, D_2)$ be the closure of all (R_1, R_2) satisfying (4)–(6) for some $(U_1, T_1, U_2, T_2, X_1, X_2, Y) \in \mathcal{P}_{D_1,D_2}$. Then

$$\mathcal{R}^*_{in}(D_1, D_2) \subseteq \mathcal{R}(D_1, D_2) \subseteq \mathcal{R}^*_{out}(D_1 + \delta, D_2 + \delta)$$

for all $\delta > 0$.

The proof is given in Appendix B.

Remark 5 The cardinality of the alphabets of the auxiliary RVs T_1 and T_2 for $\mathcal{R}^*_{in}(D_1, D_2)$ can be bounded as $|\mathcal{T}_i| \leq |\mathcal{U}_i||\mathcal{X}_i| + 1$, $i = 1, 2$; see Appendix C.

Remark 6 In the simple case of independent covertexts $Q_{U_1U_2} = Q_{U_1}Q_{U_2}$ and parallel MAAC $W_{Y|X_1X_2} = W_{Y_1|X_1}W_{Y_2|X_2}$ (where $\mathcal{Y} = \mathcal{Y}_2 \times \mathcal{Y}_2$), the inner and outer bounds of Theorem 3 coincide and reduce to the capacity formula of two parallel single-user watermarking systems [9], [15].

Example Let the covertexts be independent binary sources with $\mathcal{U}_1 = \mathcal{U}_2 = \{0,1\}$ and $Q_{U_1}(U_1 = 0) = 0.05$ and $Q_{U_2}(U_2 = 0) = 0.1$. Let the MAAC be a binary additive channel with $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{Y} = \mathcal{Z} = \{0,1\}$ and $Y = X_1 \oplus X_2 \oplus Z$, where Z is independent of (X_1, X_2) with $\Pr(Z = 1) = 0.02$ and $\oplus$ denotes modulo 2 addition. Let $D_1 = 0.45$ and $D_2 = 0.4$. Fig. 2 illustrates the numerically computed inner and outer regions of Theorems 1 and 2 (which coincide with the regions of Theorem 3 since U_1 and U_2 are independent). To compute $\mathcal{R}^*_{in}(0.45, 0.4)$, we only need to consider auxiliary RVs with alphabets $|\mathcal{T}_1| = |\mathcal{T}_2| = 5$. For comparison, we also plot two subsets of the region $\mathcal{R}^*_{out}(0.45, 0.4)$ by setting $|\mathcal{T}_1| = |\mathcal{T}_2| = 6$ and $|\mathcal{T}_1| = |\mathcal{T}_2| = 7$, respectively (recall that Theorem 3 does not give an upper bound on the alphabet sizes for T_2 and T_2 for the outer bound). It is seen that there exist noticeable gaps between $\mathcal{R}^*_{in}(0.45, 0.4)$ and the numerically obtained subsets of $\mathcal{R}^*_{out}(0.45, 0.4)$. When computing the above regions, we quantized the unit interval using a step-size of resolution 0.1 to calculate the joint distributions. We can conclude that the obtained inner and outer bounds do not coincide, and furthermore, that in case there exists a finite upper bound on the auxiliary RV alphabet sizes for the outer region, this upper bound must be at least 7 for the binary problem.

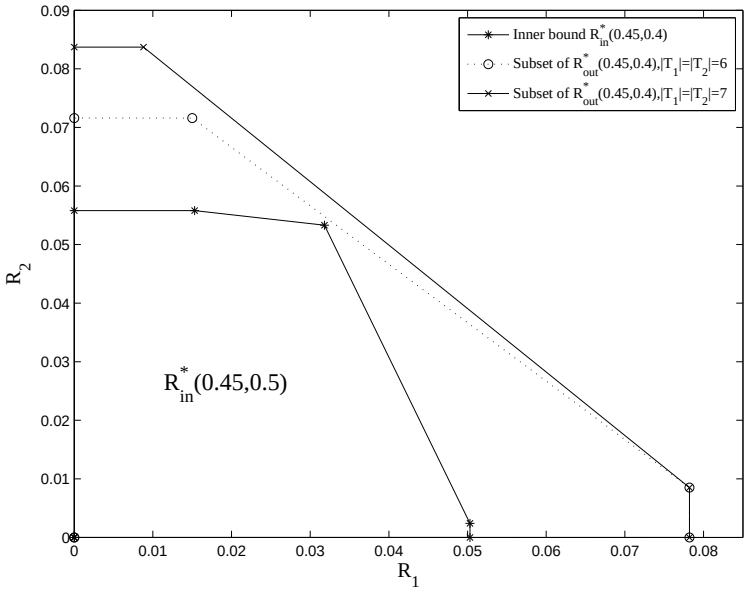

Figure 2: The inner bound $\mathcal{R}^*_{in}(0.45, 0.4)$ for the Example and two subsets of $\mathcal{R}^*_{out}(0.45, 0.4)$ obtained by setting $|\mathcal{T}_1| = |\mathcal{T}_2| = 6$ and $|\mathcal{T}_1| = |\mathcal{T}_2| = 7$. The obtained regions lie between the corresponding solid or dashed lines and the horizontal and vertical axes.

3 Proof of Theorem 1

We first recall some notation and facts regarding strongly ϵ-typicality. Let $V \triangleq (X_1, X_2, ..., X_m)$ be a superletter (a collection of RVs) taking values in a finite set $\mathcal{V} \triangleq \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_m$ and having joint distribution $P_V(x_1, ..., x_m)$, which for simplicity we also denote by $P_V(v)$. Denote by $T^{(n)}_\epsilon(V)$ or $T^{(n)}_\epsilon$ the set of all strongly ϵ-typical sequences [4, p. 326] with respect to the joint distribution $P_V(v)$. Let

$I_V \triangleq \{1,2,...,m\}$, and $I_G \subseteq I_V$. We then let $G = (X_{g_1}, X_{g_2}, ..., X_{g_{|I_G|}}) \in \mathcal{G}$ be a "sub-superletter" corresponding to I_G such that $g_i \in I_G$. Let G, K, and L be sub-superletters of V such that I_G, I_K, I_L are disjoint, and let P_G, P_K and $P_{G|K}$ be the marginal and conditional distributions induced by P_V, respectively. Denote by $T_\epsilon^{(n)}(G)$ the projection of $T_\epsilon^{(n)}(V)$ to the coordinates of G. Given any $\mathbf{k} \in \mathcal{K}^n$, denote $T_\epsilon^{(n)}(G|\mathbf{k}) \triangleq \Big\{(G^n, \mathbf{k}) \in T_\epsilon^{(n)}(G,K)\Big\}$. Clearly $T_\epsilon^{(n)}(G|\mathbf{k}) = \emptyset$ if $\mathbf{k} \notin T_\epsilon^{(n)}(K)$. The following lemma (see, e.g., [4, pp. 342–343]) restates the well known exponential bounds for the cardinality of strongly typical sets. In the lemma $\eta = \eta(\epsilon, n)$ is a generic positive term such that $\lim_{\epsilon\to 0}\lim_{n\to\infty}\eta(\epsilon,n) = 0$.

Lemma 1 [4]

1) For any $0 < \epsilon_0 < 1$ we have $P_{G|K}^{(n)}(T_\epsilon^{(n)}(G|\mathbf{k})|\mathbf{k}) > 1 - \epsilon_0$ for n sufficiently large.

2) $2^{n(H(K)-\eta)} \le \Big|\mathcal{T}_\epsilon^{(n)}(K)\Big| \le 2^{n(H(K)+\eta)}$.

3) For any $\mathbf{k} \in \mathcal{T}_\epsilon^{(n)}(K)$, $2^{n(H(G|K)-\eta)} \le \Big|\mathcal{T}_\epsilon^{(n)}(G|\mathbf{k})\Big| \le 2^{n(H(G|K)+\eta)}$.

Finally, we recall the Markov lemma for joint strong ϵ-typicality.

Lemma 2 (Markov lemma [4, p. 579]) Let $G \to K \to L$ form a Markov chain in this order. For any $0 < \epsilon_0 < 1$ and $(\mathbf{g}, \mathbf{k}) \in T_\epsilon^{(n)}(G,K)$,

$$P_{L|K}^{(n)}\Big((\mathbf{g},\mathbf{k},L^n) \in T_\epsilon^{(n)}(G,K,L)\Big|\,\mathbf{k}\Big) > 1-\epsilon_0$$

for n sufficiently large, independently of $(\mathbf{g}, \mathbf{k})$.

3.1 Outline of Proof

It is enough to show that for given $Q_{U_1U_2}$, $W_{Y|X_1X_2}$, and any $(R_1, R_2) \in \mathcal{R}_{in}(D_1, D_2)$, there exists a sequence of codes $(f_1^{(n)}, f_2^{(n)}, \psi^{(n)})$ such that $P_e^{(n)} \to 0$ as $n \to \infty$ and for any $\delta > 0$,

$$\frac{1}{n}\mathbb{E}[d_i(U_i^n, f_i^{(n)}(W_i, U_i^n))] \le D_i + \delta, \quad i = 1,2$$

for n sufficiently large. Once this is proved, a standard subsequence diagonalization argument can be used to prove a similar statement with $\delta = 0$, which then directly implies the theorem.

Fix $(P_{T_1|U_1}, P_{X_1|U_1T_1}, P_{T_2|U_2}, P_{X_2|U_2T_2})$ such that $I(U_i;T_i) > 0$ and the following are satisfied for some $\epsilon' > 0$,

$$R_1 < I(T_1;T_2,Y) - I(U_1;T_1) - \epsilon', \tag{7}$$
$$R_2 < I(T_2;T_1,Y) - I(U_2;T_2) - \epsilon', \tag{8}$$
$$R_1 + R_2 < I(T_1,T_2;Y) - I(U_1,U_2;T_1,T_2) - \epsilon', \tag{9}$$
$$\mathbb{E}[d_i(U_i,X_i)] \le D_i,\ i = 1,2. \tag{10}$$

We will choose $f_1^{(n)}$ and $f_2^{(n)}$ in a random manner. For $\epsilon < \frac{\delta}{2\max\{d_1^{max}, d_2^{max}\}}$, define

$$P_i^{(n)} \triangleq \Pr\Big(\frac{1}{n}d_i\big(U_i^n, f_i^{(n)}(W_i,U_i^n)\big) > D_i + \epsilon d_i^{max}\Big),\ i = 1,2.$$

The technically challenging part of the proof is to show that for any $0 < \epsilon_1 \leq \frac{\delta}{6\max\{d_1^{max}, d_2^{max}\}}$, the probabilities $P_e^{(n)}$, $P_1^{(n)}$, and $P_2^{(n)}$, when averaged over the random choice of $f_1^{(n)}$ and $f_2^{(n)}$, satisfy

$$\mathbb{E}[P_e^{(n)}] \leq \epsilon_1, \quad \mathbb{E}[P_1^{(n)}] \leq \epsilon_1, \quad \mathbb{E}[P_2^{(n)}] \leq \epsilon_1$$

for n sufficiently large. Then $\mathbb{E}\{P_e^{(n)} + P_1^{(n)} + P_2^{(n)}\} \leq 3\epsilon_1$, which guarantees that there exists at least one pair of codes $(f_1^{(n)}, f_2^{(n)})$ such that $P_e^{(n)} + P_1^{(n)} + P_2^{(n)} \leq 3\epsilon_1$ and hence $P_e^{(n)} \leq 3\epsilon_1$, $P_1^{(n)} \leq 3\epsilon_1$, $P_2^{(n)} \leq 3\epsilon_1$ are simultaneously satisfied for n sufficiently large. Finally, it can be easily shown that $P_i^{(n)} \leq 3\epsilon_1$ implies for n sufficiently large that

$$\frac{1}{n}\mathbb{E}\left[d_i(U_i^n, f_i^{(n)}(W_i, U_i^n)\right] \leq D_i + \epsilon d_i^{max} + P_i^{(n)} d_i^{max} \leq D_i + \delta.$$

3.2 Random Code Design

In what follows, the strongly ϵ-typical set $\mathcal{T}_\epsilon^{(n)}$ is defined under the joint distribution

$$P_{U_1U_2T_1T_2X_1X_2Y} = Q_{U_1U_2} P_{T_1|U_1} P_{X_1|U_1T_1} P_{T_2|U_2} P_{X_2|U_2T_2} W_{Y|X_1X_2} \tag{11}$$

and all the marginal and conditional distributions, e.g., $P_{U_2T_2}$, $P_{U_1|U_2T_2}$, etc, are induced by the joint distribution. The parameter ϵ, which is chosen to be sufficiently small, will be specified in the proof.

Generation of codebooks. For $i = 1, 2$ and every $w_i \in \mathcal{W}_i$, generate a codebook

$$\mathcal{C}_{w_i} = \{\mathbf{t}_i(w_i, 1), \mathbf{t}_i(w_i, 2), ..., \mathbf{t}_i(w_i, L_i)\}$$

with $L_i = 2^{n[I(U_i;T_i)+4\epsilon]}$ codewords such that each $\mathbf{t}_i(w_i, l_i)$ is independently selected with uniform distribution from the typical set $\mathcal{T}_\epsilon^{(n)}(T_i)$. Denote the entire codebook for Encoder i by $\mathcal{C}^{(i)} = \{\mathcal{C}_{w_i}\}_{w_i=1}^{M_i}$, where we recall that $M_i = 2^{nR_i}$. For each $\mathbf{u}_i$ and codeword $\mathbf{t}_i(w_i, l_i)$ $(1 \leq w_i \leq M_i, 1 \leq l_i \leq L_i)$, generate a codeword $\mathbf{x}_i$ according to $P_{X_i|U_iT_i}^{(n)}(\mathbf{x}_i|\mathbf{u}_i, \mathbf{t}_i)$. Denote the codebook of all the codewords $\mathbf{x}_i$ by $\mathcal{B}^{(i)}$.

Encoder $f_1^{(n)}$: Encoder $f_1^{(n)}$ is the concatenation of a pre-encoder $\varphi_1^{(n)} : \mathcal{W}_1 \times \mathcal{U}_1^n \longrightarrow \mathcal{T}_1^n$ and a mapping $g_1^{(n)} : \mathcal{U}_1^n \times \mathcal{T}_1^n \longrightarrow \mathcal{X}_1^n$.

To define $\varphi_1^{(n)}$, we need the following notation adopted from [11]. We introduce a conditional probability

$$A^{(n)}(\mathbf{u}_1, \mathbf{t}_1) \triangleq P_{U_2T_2|U_1T_1}^{(n)}\left((\mathbf{u}_2, \mathbf{t}_2) : (\mathbf{u}_2, \mathbf{t}_2) \in \mathcal{T}_\epsilon^{(n)}(U_2T_2|\mathbf{u}_1, \mathbf{t}_1)\middle|\, \mathbf{u}_1, \mathbf{t}_1\right).$$

For $\mu \in (0, 1)$, let

$$\mathcal{F}_{\mu,\epsilon}^{(n)}(U_1, T_1) \triangleq \left\{(\mathbf{u}_1, \mathbf{t}_1) : A^{(n)}(\mathbf{u}_1, \mathbf{t}_1) \geq 1 - \mu\right\}.$$

By definition, we have $\mathcal{F}_{\mu,\epsilon}^{(n)}(U_1, T_1) \subseteq \mathcal{T}_\epsilon^{(n)}(U_1, T_1)$.

We now describe the pre-encoding function $\varphi_1^{(n)} = \varphi_1^{(n)}(w_1, \mathbf{u}_1)$ which maps every pair $(w_1, \mathbf{u}_1)$ to a codeword in $\mathcal{C}^{(1)} \subseteq \mathcal{T}_1^n$. Given $w_1 \in \{1, 2, ..., M_1\}$ and $\mathbf{u}_1$, $\varphi_1^{(n)}$ seeks the first codeword $\mathbf{t}_1(w_1, l_1)$ (if any) in $\mathcal{C}_{w_1}$ such that $(\mathbf{u}_1, \mathbf{t}_1(w_1, l_1)) \in \mathcal{F}_{\mu,\epsilon}^{(n)}(U_1, T_1)$. If there is no such codeword, $\varphi_1^{(n)}$ outputs $\mathbf{t}_1(w_1, 1)$. Next, for each output $\mathbf{t}_1(w_1, l_1)$ and $\mathbf{u}_1$, $g_1^{(n)}$ sends out the associated codeword $\mathbf{x}_1(w_1, \mathbf{u}_1)$ to the channel. Thus, $f_1^{(n)}(w_1, \mathbf{u}_1) = g_1^{(n)}\left(\mathbf{u}_1, \varphi_1^{(n)}(w_1, \mathbf{u}_1)\right)$.

Encoder $f_2^{(n)}$: Encoder $f_2^{(n)}$ is the concatenation of a pre-encoder $\varphi_2^{(n)} : \mathcal{W}_2 \times \mathcal{U}_2^n \longrightarrow \mathcal{T}_2^n$ and a mapping $g_2^{(n)} : \mathcal{U}_2^n \times \mathcal{T}_2^n \longrightarrow \mathcal{X}_2^n$.

To define $\varphi_2^{(n)}$, let

$$B_{\varphi_1}^{(n)}(\mathbf{u}_2, \mathbf{t}_2) \triangleq \frac{1}{2^{nR_1}} \sum_{w_1=1}^{M_1} P_{U_1|U_2T_2}^{(n)} \left(\mathbf{u}_1 : (\mathbf{u}_1, \varphi_1^{(n)}(w_1, \mathbf{u}_1)) \in \mathcal{T}_\epsilon^{(n)}(U_1T_1|\mathbf{u}_2, \mathbf{t}_2) \middle| \mathbf{u}_2, \mathbf{t}_2 \right).$$

Also, for $\nu \in (0, 1)$, define

$$\mathcal{F}_{\varphi_1,\nu,\epsilon}^{(n)}(U_2, T_2) \triangleq \left\{ (\mathbf{u}_2, \mathbf{t}_2) : B_{\varphi_1}^{(n)}(\mathbf{u}_2, \mathbf{t}_2) \geq 1 - \nu \right\}.$$

By definition, it is seen that $\mathcal{F}_{\varphi_1,\nu,\epsilon}^{(n)}(U_2, T_2) \subseteq \mathcal{T}_\epsilon^{(n)}(U_2, T_2)$.

We now describe the pre-encoding function $\varphi_2^{(n)} = \varphi_2^{(n)}(w_2, \mathbf{u}_2)$ which maps every pair $(w_2, \mathbf{u}_2)$ to a codeword in $\mathcal{C}^{(2)} \subseteq \mathcal{T}_2^n$. Given $w_2 \in \{1, 2, ..., M_2\}$ and $\mathbf{u}_2$, $\varphi_2^{(n)}$ seeks the first codeword $\mathbf{t}_2(w_2, l_2)$ (if any) in $\mathcal{C}_{w_2}$ such that $(\mathbf{u}_2, \mathbf{t}_2(w_2, l_2)) \in \mathcal{F}_{\varphi_1,\nu,\epsilon}^{(n)}(U_2, T_2)$. If there is no such codeword, $\varphi_2^{(n)}$ outputs $\mathbf{t}_2(w_2, 1)$. Next, for each output $\mathbf{t}_2(w_2, l_2)$, $g_2^{(n)}$ sends out the associated codeword $\mathbf{x}_2(w_2, \mathbf{u}_2)$ to the channel. Thus, $f_2^{(n)}(w_2, \mathbf{u}_2) = g_2^{(n)} \left(\mathbf{u}_2, \varphi_2^{(n)}(w_2, \mathbf{u}_2) \right)$.

Decoder $\psi^{(n)}$: Given $\mathbf{y}$, $\psi^{(n)}$ seeks $\mathbf{t}_1(\widehat{w}_1, \widehat{l}_1) \in \mathcal{C}^{(1)}$ and $\mathbf{t}_2(\widehat{w}_2, \widehat{l}_2) \in \mathcal{C}^{(2)}$ such that

$$(\mathbf{t}_1(\widehat{w}_1, \widehat{l}_1), \mathbf{t}_2(\widehat{w}_2, \widehat{l}_2), \mathbf{y}) \in \mathcal{T}_\epsilon^{(n)}(T_1, T_2, Y).$$

If such a pair $(\mathbf{t}_1(\widehat{w}_1, \widehat{l}_1), \mathbf{t}_2(\widehat{w}_2, \widehat{l}_2))$ exists for a unique $(\widehat{w}_1, \widehat{w}_2)$, then $\psi^{(n)}$ outputs $\widehat{w}_1$ and $\widehat{w}_2$ as the decoded messages. If there is no such pair $(\widehat{w}_1, \widehat{w}_2)$, or it is not unique, a decoding error is declared. Letting $\mathbf{t}_i(w_i, l_i) = \varphi_i^{(n)}(w_i, \mathbf{u}_i)$, it is easy to see that if there is a decoding error, then at least one of the following events occurs:

1) E_1: $(\mathbf{t}_1(w_1, l_1), \mathbf{t}_2(w_2, l_2), \mathbf{y}) \notin \mathcal{T}_\epsilon^{(n)}(T_1, T_2, Y)$,

2) E_2: there exist l_1' and $w_1' \neq w_1$ and l_2' (l_2' may or may not be equal to l_2) such that

$$(\mathbf{t}_1(w_1', l_1'), \mathbf{t}_2(w_2, l_2'), \mathbf{y}) \in \mathcal{T}_\epsilon^{(n)}(T_1, T_2, Y),$$

3) E_3: there exist l_2' and $w_2' \neq w_2$ and l_1' (l_1' may or may not be equal to l_1) such that

$$(\mathbf{t}_1(w_1, l_1'), \mathbf{t}_2(w_2', l_2'), \mathbf{y}) \in \mathcal{T}_\epsilon^{(n)}(T_1, T_2, Y),$$

or

4) E_4: there exist l_1' and $w_1' \neq w_1$ and l_2' and $w_2' \neq w_2$ such that

$$(\mathbf{t}_1(w_1', l_1'), \mathbf{t}_2(w_2', l_2'), \mathbf{y}) \in \mathcal{T}_\epsilon^{(n)}(T_1, T_2, Y).$$

In the following, we will bound the probabilities $P_e^{(n)}$, $P_1^{(n)}$ and $P_2^{(n)}$ averaged over the random choice of all codes $\mathcal{B}^{(1)}$, $\mathcal{B}^{(2)}$, $\mathcal{C}^{(1)}$, and $\mathcal{C}^{(2)}$. To simplify the notation we abbreviate $\mathbb{E}_{\mathcal{B}^{(1)},\mathcal{B}^{(2)},\mathcal{C}^{(1)},\mathcal{C}^{(2)}}[\,\cdot\,]$ as $\mathbb{E}_\Omega[\,\cdot\,]$.

3.3 Bounding $\mathbb{E}_\Omega[P_e^{(n)}]$

To analyze the average probability of error, we need the following lemmas.

Lemma 3 For any $w_1 \in \mathcal{W}_1$, $w_2 \in \mathcal{W}_2$, and any $\epsilon_0, \epsilon \in (0,1)$, one can choose $\mu, \nu \in (0,1)$ small enough such that

$$\mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}\left[P^{(n)}_{U_1U_2}\left((\varphi_1^{(n)}(w_1,\mathbf{u}_1),\mathbf{u}_1,\mathbf{u}_2,\varphi_2^{(n)}(w_2,\mathbf{u}_2)) \in \mathcal{T}_\epsilon^{(n)}(T_1,U_1,U_2,T_2)\right)\right] \geq 1-\epsilon_0$$

for n sufficiently large, where the expectation is taken with respect to the random codes $\mathcal{C}^{(1)}$ and $\mathcal{C}^{(2)}$.

The proof of Lemma 3 is very similar to the proof of the extended Markov lemma in [11, Lemma 3] for correlated Gaussian sources and is hence omitted; readers may also refer to [14, Section 5.4.5].

Since the watermarks are independently and uniformly distributed, and by the symmetry of the code construction, we can assume without the loss of generality that some fixed $w_1 \in \mathcal{W}_1$ and $w_2 \in \mathcal{W}_2$ are the transmitted watermarks. Thus we bound the probability of error as

$$\begin{aligned} P_e^{(n)} &= \Pr\left(\left\{\psi^{(n)}(Y^n) \neq (w_1,w_2)\right\}\right) \\ &\leq \Pr(A_1) + \Pr\left(\left\{\psi^{(n)}(Y^n) \neq (w_1,w_2)\right\}\middle| A_1^c\right) \end{aligned} \tag{12}$$

where A_1 is the event

$$A_1 : (\mathbf{t}_1(w_1,l_1),\mathbf{u}_1,\mathbf{u}_2,\mathbf{t}_2(w_2,l_2),\mathbf{x}_1,\mathbf{x}_2) \notin \mathcal{T}_\epsilon^{(n)}(T_1,U_1,U_2,T_2,X_1,X_2).$$

Recall that $\mathbf{t}_i(w_i,l_i) = \varphi_i^{(n)}(w_i,\mathbf{u}_i)$, $i=1,2$. We also let $\mathbf{t}_i(w_i,l_i')$ and $\mathbf{t}_i(w_i',l_i')$ be the l_i'-th codeword in the codebook $\mathcal{C}_{w_i}$ and $\mathcal{C}_{w_i'}$, respectively.

We then introduce the event

$$A_0 : (\mathbf{t}_1(w_1,l_1),\mathbf{u}_1,\mathbf{u}_2,\mathbf{t}_2(w_2,l_2)) \notin \mathcal{T}_\epsilon^{(n)}(T_1,U_1,U_2,T_2).$$

Taking expectation in (12) and using the union bound, we have

$$\mathbb{E}_\Omega[P_e^{(n)}] \leq \mathbb{E}_\Omega \Pr(A_0) + \mathbb{E}_\Omega \Pr(A_1|A_0^c) + \mathbb{E}_\Omega \Pr(E_1|A_1^c) + \sum_{k=2}^{4} \mathbb{E}_\Omega \Pr(E_k|A_1^c). \tag{13}$$

It immediately follows from Lemma 3 that

$$\mathbb{E}_\Omega \Pr(A_0) = \mathbb{E}_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}} \Pr(A_0) \leq \epsilon_0 \tag{14}$$

for n sufficiently large, where we set $\epsilon_0 = \epsilon_1/7$ for a given $\epsilon_1 \geq 0$ throughout the proof. When A_0^c holds, since $\mathbf{x}_1$ and $\mathbf{x}_2$ are respectively drawn according to the conditional probabilities $P^{(n)}_{X_1|U_1T_1}(\cdot|\mathbf{u}_1,\mathbf{t}_1)$ and $P^{(n)}_{X_2|U_2T_2}(\cdot|\mathbf{u}_2,\mathbf{t}_2)$, and $\mathbf{y}$ is drawn according to the conditional distribution $W^{(n)}_{Y|X_1X_2}(\cdot|\mathbf{x}_1,\mathbf{x}_2)$, it follows from two successive applications of Lemma 2 that

$$\mathbb{E}_\Omega \Pr(A_1|A_0^c) \leq \mathbb{E}_\Omega[\epsilon_0] = \epsilon_0 \tag{15}$$

and

$$\begin{aligned}
&\mathbb{E}_\Omega \Pr\left(E_1 \middle| A_1^c\right) \\
&\quad\leq \mathbb{E}_\Omega \Pr\left(\left\{\left(\varphi_1^{(n)}(w_1,U_1^n),U_1^n,U_2^n,\varphi_2^{(n)}(w_2,U_2^n),f_1^{(n)}(w_1,U_1^n),f_2^{(n)}(w_2,U_2^n),Y^n\right)\notin\mathcal{T}_\epsilon^{(n)}\right\}\middle| A_1^c\right) \\
&\quad\leq \mathbb{E}_\Omega[\epsilon_0]=\epsilon_0 \qquad (16)
\end{aligned}$$

for n sufficiently large. It remains to bound $\mathbb{E}_\Omega \Pr\{E_k|A_1^c\}$ for $k=2,3,4$. Using the union bound we write

$$\begin{aligned}
&\mathbb{E}_\Omega \Pr\left(E_2\middle| A_1^c\right) \\
&\quad\leq \sum_{w_1'\neq w_1}\sum_{l_1'=1}^{L_1}\Pr\left(\left\{(T_1^n(w_1',l_1'),Y^n,T_2^n(w_2,l_2'))\in\mathcal{T}_\epsilon^{(n)}(T_1,T_2,Y)\right\}\middle| A_1^c\right), \qquad (17)
\end{aligned}$$

where $T_1^n(w_1',l_1')$ is a RV uniformly drawn from $\mathcal{T}_\epsilon^{(n)}(T_1)$ which is independent of $(T_2^n(w_2,l_2'),Y^n)$ since $w_1'\neq w_1$. Thus we have

$$\begin{aligned}
&\Pr\left(\left\{(T_1^n(w_1',l_1'),Y^n,T_2^n(w_2,l_2'))\in\mathcal{T}_\epsilon^{(n)}(T_1,T_2,Y)\right\}\middle| A_1^c\right) \\
&\quad= \sum_{(\mathbf{t}_2,\mathbf{y})\in\mathcal{T}_\epsilon^{(n)}(T_2,Y)}\ \sum_{\mathbf{t}_1\in\mathcal{T}_\epsilon^{(n)}(T_1|\mathbf{t}_2,\mathbf{y})}\Pr\left(T_2^n(w_2,l_2')=\mathbf{t}_2,Y^n=\mathbf{y}\middle| A_1^c\right) \\
&\qquad\qquad \Pr\left(T_1^n(w_1',l_1')=\mathbf{t}_1\middle| T_2^n(w_2,l_2')=\mathbf{t}_2,Y^n=\mathbf{y},A_1^c\right) \\
&\quad= \sum_{(\mathbf{t}_2,\mathbf{y})\in\mathcal{T}_\epsilon^{(n)}(T_2,Y)}\ \sum_{\mathbf{t}_1\in\mathcal{T}_\epsilon^{(n)}(T_1|\mathbf{t}_2,\mathbf{y})}\Pr\left(T_2^n(w_2,l_2')=\mathbf{t}_2,Y^n=\mathbf{y}\middle| A_1^c\right)\Pr\left(T_1^n(w_1',l_1')=\mathbf{t}_1\right) \\
&\quad= \sum_{(\mathbf{t}_2,\mathbf{y})\in\mathcal{T}_\epsilon^{(n)}(T_2,Y)}\Pr\left(T_2^n(w_2,l_2)=\mathbf{t}_2,Y^n=\mathbf{y}\middle| A_1^c\right)\frac{|\mathcal{T}_\epsilon^{(n)}(T_1|\mathbf{t}_2,\mathbf{y})|}{|\mathcal{T}_\epsilon^{(n)}(T_1)|} \\
&\quad\leq \frac{2^{n[H(T_1|T_2,Y)+\eta]}}{2^{n[H(T_1)-\eta]}}\sum_{(\mathbf{t}_2,\mathbf{y})\in\mathcal{T}_\epsilon^{(n)}(T_2,Y)}\Pr\left(T_2^n(w_2,l_2')=\mathbf{t}_2,Y^n=\mathbf{y}\middle| A_1^c\right) \\
&\quad\leq 2^{-n[I(T_1;T_2,Y)-2\eta]}, \qquad (18)
\end{aligned}$$

where the first inequality follows from Lemma 1. Recalling that $\eta\to 0$ as $n\to\infty$ and $\epsilon\to 0$, we can make sure that $2\eta<\epsilon'-4\epsilon$ by choosing ϵ small enough and n large enough. Thus from (17)

$$\begin{aligned}
\mathbb{E}_\Omega \Pr\left(E_2\middle| A_1^c\right) &\leq 2^{n[R_1+I(U_1;T_1)+4\epsilon-I(T_1;T_2,Y)+2\eta]} \\
&\leq 2^{n[R_1+I(U_1;T_1)-I(T_1;T_2,Y)+\epsilon']} \\
&\leq \epsilon_0 \qquad (19)
\end{aligned}$$

for ϵ sufficiently small and n sufficiently large, where (19) follows from the assumption (7). Similarly we have

$$\mathbb{E}_\Omega \Pr\left(E_3\middle| A_1^c\right)\leq\epsilon_0 \qquad (20)$$

for ϵ small enough and n sufficiently large. We next bound

$$\begin{aligned}
&\mathbb{E}_\Omega \Pr\left(E_4\middle| A_1^c\right) \\
&\quad\leq \sum_{w_1'\neq w_1}\sum_{l_1'=1}^{L_1}\sum_{w_2'\neq w_2}\sum_{l_2'=1}^{L_2}\Pr\left(\left\{(T_1^n(w_1',l_1'),T_2^n(w_2',l_2'),Y^n)\in\mathcal{T}_\epsilon^{(n)}(T_1,T_2,Y)\right\}\middle| A_1^c\right),
\end{aligned}$$

where $T_1^n(w_1', l_1')$ and $T_2^n(w_2', l_2')$ are RVs independently drawn from $\mathcal{T}_\epsilon^{(n)}(T_1)$ and $\mathcal{T}_\epsilon^{(n)}(T_2)$ according to the uniform distribution, respectively. We have

$$\begin{aligned}
&\Pr\Big(\Big\{(T_1^n(w_1', l_1'), T_2^n(w_2', l_2'), Y^n) \in \mathcal{T}_\epsilon^{(n)}(T_1, T_2, Y)\Big\}\Big| A_1^c\Big) \\
&= \sum_{\mathbf{y}\in\mathcal{T}_\epsilon^{(n)}(Y)} \sum_{(\mathbf{t}_1,\mathbf{t}_2)\in\mathcal{T}_\epsilon^{(n)}(T_1,T_2|\mathbf{y})} \Pr(Y^n = \mathbf{y}|A_1^c) \\
&\qquad \Pr(T_1^n(w_1', l_1') = \mathbf{t}_1, T_2^n(w_2', l_2') = \mathbf{t}_2|A_1^c, Y^n = \mathbf{y}) \\
&= \sum_{\mathbf{y}\in\mathcal{T}_\epsilon^{(n)}(Y)} \sum_{(\mathbf{t}_1,\mathbf{t}_2)\in\mathcal{T}_\epsilon^{(n)}(T_1,T_2|Y)} \Pr(Y^n = \mathbf{y}|A_1^c) \frac{1}{|\mathcal{T}_\epsilon^{(n)}(T_1)|} \frac{1}{|\mathcal{T}_\epsilon^{(n)}(T_2)|} \\
&\le \sum_{\mathbf{y}\in\mathcal{T}_\epsilon^{(n)}(Y)} \Pr(Y^n = \mathbf{y}|A_1^c) \frac{2^{n[H(T_1,T_2|Y)+\eta]}}{2^{n[H(T_1)-\eta]}2^{n[H(T_2)-\eta]}} \\
&\le 2^{-n[I(T_1,T_2;Y)+I(T_1;T_2)-3\eta]}
\end{aligned}$$

and hence

$$\begin{aligned}
&\mathbb{E}_\Omega \Pr\left(E_4|\, A_1^c\right) \\
&\quad\le 2^{n[R_1+R_2+I(U_1;T_1)+I(U_2;T_2)-I(T_1,T_2;Y)-I(T_1;T_2)+8\epsilon+3\eta]} \\
&\quad\le 2^{n[R_1+I(U_1,U_2;T_1,T_2)-I(T_1,T_2;Y)+\epsilon']} \\
&\quad\le \epsilon_0
\end{aligned} \tag{21}$$

for n sufficiently large and ϵ small enough (such that $8\epsilon+3\eta<\epsilon'$), where the second inequality holds by the Markov chain relation $T_1 \to U_1 \to U_2 \to T_2$ imposed in Definition 2, and the last inequality follows from the assumption (9). Finally, substituting (14)–(16), (19), (20) and (21) into (13) yields $\mathbb{E}_\Omega[P_e^{(n)}] \le 7\epsilon_0 = \epsilon_1$ for ϵ sufficiently small and n sufficiently large.

3.4 Bounding $\mathbb{E}_\Omega[P_i^{(n)}]$

We only bound $\mathbb{E}_\Omega[P_i^{(n)}]$ for $i=1$, since the case $i=2$ can be dealt with similarly. When $(\mathbf{u}_1, \mathbf{x}_1(w_1, \mathbf{u}_1)) \in \mathcal{T}_\epsilon^{(n)}(U_1, X_1)$,

$$\frac{1}{n} d_1\big(\mathbf{u}_1, \mathbf{x}_1(w_1, \mathbf{u}_1)\big) \le \mathbb{E}[d_1(U_1, X_1)] + \epsilon d_1^{max} \le D_1 + \epsilon d_1^{max}$$

for n sufficiently large, where the first inequality follows from the definition of strong typicality and the second inequality follows from (10). This means that if $\frac{1}{n} d_1\big(U_1^n, f_1^{(n)}(W_1, U_1^n)\big) > D_1 + \epsilon d_1^{max}$, then we must have $\big(U_1^n, f_1^{(n)}(W_1, U_1^n)\big) \notin \mathcal{T}_\epsilon^{(n)}(U_1, X_1)$ for n sufficiently large. Thus, we can bound

$$\begin{aligned}
&\Pr\Big(\frac{1}{n} d_1(U_1^n, f_1^{(n)}(W_1, U_1^n)) > D_1 + \epsilon d_1^{max}\Big) \\
&\quad\le \Pr\Big(\big(U_1^n, f_1^{(n)}(W_1, U_1^n)\big) \notin \mathcal{T}_\epsilon^{(n)}(U_1, X_1)\Big) \\
&\quad\le \Pr\Big(\big(U_1^n, \varphi_1^{(n)}(W_1, U_1^n), f_1^{(n)}(W_1, U_1^n)\big) \notin \mathcal{T}_\epsilon^{(n)}(U_1, T_1, X_1)\Big) \\
&\quad\le \Pr\Big(\big(U_1^n, \varphi_1^{(n)}(W_1, U_1^n)\big) \notin \mathcal{T}_\epsilon^{(n)}(U_1, T_1)\Big) \\
&\qquad + \Pr\Big(\big(U_1^n, \varphi_1^{(n)}(W_1, U_1^n), f_1^{(n)}(W_1, U_1^n)\big) \notin \mathcal{T}_\epsilon^{(n)}(U_1, T_1, X_1)\Big| \big(U_1^n, \varphi_1^{(n)}(W_1, U_1^n)\big) \in \mathcal{T}_\epsilon^{(n)}(U_1, T_1)\Big)
\end{aligned}$$

$$\leq \Pr\left(\left(\varphi_1^{(n)}(W_1,U_1^n),U_1^n,U_2^n,\varphi_2^{(n)}(W_2,U_2^n)\right)\notin\mathcal{T}_\epsilon^{(n)}(T_1,U_1,U_2,T_2)\right)$$
$$+\Pr\left(\left(U_1^n,\varphi_1^{(n)}(W_1,U_1^n),f_1^{(n)}(W_1,U_1^n)\right)\notin\mathcal{T}_\epsilon^{(n)}(U_1,T_1,X_1)\middle|\,(U_1^n,\varphi_1^{(n)}(W_1,U_1^n))\in\mathcal{T}_\epsilon^{(n)}(U_1,T_1)\right). \tag{22}$$

Now taking expectation on both sides, the first term of (22) is bounded by $\frac{\epsilon_1}{2}$ by Lemma 3, and the second term is bounded by $\frac{\epsilon_1}{2}$ for sufficiently large n by Lemma 1. This completes the proof of the bound $\mathbb{E}_\Omega[P_1^{(n)}]\leq\epsilon_1$ for n sufficiently large. □

4 Concluding Remarks

We have studied a multi-user information embedding system consisting of two information embedders and one joint decoder connected via a multiple-access attack channel. We have obtained an inner bound for the capacity region in a computable single-letter form. We also derived an outer bound for the capacity region, but in this case the auxiliary random variables involved in the region's characterization have no upper bounds on their alphabet's cardinality. Consequently, there may not exist an algorithm to compute the outer bound with arbitrary precision. We have also addressed the special case when the covertexts are independent of each other and inner and outer bounds for the capacity region of this simplified system are provided. Finally, we remark that using a similar technique inner and outer bounds are derived in [14, Chapter 5] for the capacity region of private multi-user embedding systems with quantization.

Appendix

A Proof of Theorem 2

The proof is a generalization of the proof of the converse in [15] for a single-user embedding system.

We need to show that any MAE code $(f_1^{(n)},f_2^{(n)},\psi^{(n)})$ with achievable rate pair (R_1,R_2) must satisfy (1)–(3) for some auxiliary RVs T_1 and T_2 with joint distribution $P_{U_1U_2T_1T_2X_1X_2Y}\in\mathcal{P}_{D_1,D_2}$. It follows from Fano's inequality that

$$H(W_1,W_2|Y^n)\leq n(R_1+R_2)P_e^{(n)}+H(P_e^{(n)})\triangleq n\epsilon_n.$$

It is clear that $\epsilon_n\to 0$ if $P_e^{(n)}\to 0$ and

$$\begin{aligned}H(W_1|Y^n)&\leq H(W_1,W_2|Y^n)\leq n\epsilon_n,\\ H(W_2|Y^n)&\leq H(W_1,W_2|Y^n)\leq n\epsilon_n.\end{aligned}$$

Because W_1 is uniformly drawn from the message set $\{1,2,...,2^{nR_1}\}$ and is independent of U_1^n, we have

$$nR_1=H(W_1)=I(W_1;Y^n)+H(W_1|Y^n)\leq I(W_1;Y^n)-\underbrace{I(W_1;U_1^n)}_{=0}+n\epsilon_n.$$

Hence we can write

$$
\begin{aligned}
& I(W_1;Y^n) - I(W_1;U_1^n) \\
& \overset{(a)}{=} \sum_{k=1}^{n} \Big[I(W_1;Y_k|Y_1^{k-1}) - I(W_1;U_{1k}|U_{1,k+1}^n) \Big] \\
& = \sum_{k=1}^{n} \Big[H(Y_k|Y_1^{k-1}) - H(Y_k|W_1,Y_1^{k-1},U_{1,k+1}^n) - I(Y_k;U_{1,k+1}^n|W_1,Y_1^{k-1}) \\
& \qquad -H(U_{1k}|U_{1,k+1}^n) + H(U_{1k}|W_1,U_{1,k+1}^n) \Big] \\
& \overset{(b)}{=} \sum_{k=1}^{n} \Big[H(Y_k|Y_1^{k-1}) - H(Y_k|W_1,Y_1^{k-1},U_{1,k+1}^n) - I(U_{1k};Y_1^{k-1}|W_1,U_{1,k+1}^n) \\
& \qquad -H(U_{1k}|U_{1,k+1}^n) + H(U_{1k}|W_1,U_{1,k+1}^n) \Big] \\
& \overset{(c)}{=} \sum_{k=1}^{n} \Big[H(Y_k|Y_1^{k-1}) - H(Y_k|W_1,Y_1^{k-1},U_{1,k+1}^n) \\
& \qquad -H(U_{1k}) + H(U_{1k}|W_1,Y_1^{k-1},U_{1,k+1}^n) \Big] \\
& \leq \sum_{k=1}^{n} \Big[H(Y_k) - H(Y_k|W_1,U_{1,k+1}^n,Y_1^{k-1}) - I(U_{1k};W_1,Y_1^{k-1},U_{1,k+1}^n) \Big] \\
& = \sum_{k=1}^{n} \Big[I(Y_k;W_1,U_{1,k+1}^n,Y_1^{k-1}) - I(U_{1k};W_1,Y_1^{k-1},U_{1,k+1}^n) \Big] \\
& \overset{(d)}{\leq} \sum_{k=1}^{n} \Big[I(W_2,U_{2,k+1}^n,Y_1^{k-1},Y_k;W_1,U_{1,k+1}^n,Y_1^{k-1}) - I(U_{1k};W_1,Y_1^{k-1},U_{1,k+1}^n) \Big] \\
& \overset{(e)}{=} \sum_{k=1}^{n} [I(L_{2k},Y_k;L_{1k}) - I(U_{1k};L_{1k})]
\end{aligned}
$$

where in (a) $Y_1^{k-1} \triangleq (Y_1, Y_2, ..., Y_{k-1})$ and $U_{1,k+1}^n \triangleq (U_{1,k+1}, U_{1,k+2}, ..., U_{1,n})$, (b) follows from the "summation by parts" identity [3, Lemma 7], (c) holds since the source U_1 is memoryless, in (d) $U_{2,k+1}^n \triangleq (U_{2,k+1}, U_{2,k+2}, ..., U_{2,n})$, and in (e) $L_{1k} \triangleq (W_1, Y_1^{k-1}, U_{1,k+1}^n)$ and $L_{2k} \triangleq (W_2, Y_1^{k-1}, U_{2,k+1}^n)$. Hence we obtain the bound

$$R_1 \leq \frac{1}{n} \sum_{k=1}^{n} [I(L_{1k};L_{2k},Y_k) - I(U_{1k};L_{1k})] + \epsilon_n. \tag{23}$$

Similarly, we can show that

$$R_2 \leq \frac{1}{n} \sum_{k=1}^{n} [I(L_{2k};L_{1k},Y_k) - I(U_{2k};L_{2k})] + \epsilon_n. \tag{24}$$

To bound the sum of the rates, we write

$$
\begin{aligned}
n(R_1+R_2) &= H(W_1,W_2) = I(W_1,W_2;Y^n) + H(W_1,W_2|Y^n) \\
&\leq I(W_1,W_2;Y^n) - \underbrace{I(W_1,W_2;U_1^n,U_2^n)}_{=0} + n\epsilon_n
\end{aligned} \tag{25}
$$

and

$$
\begin{aligned}
&I(W_1,W_2;Y^n) - I(W_1,W_2;U_1^n,U_2^n) \\
&= \sum_{k=1}^{n}\Big[I(W_1,W_2;Y_k|Y_1^{k-1}) - I(W_1,W_2;U_{1k},U_{2k}|U_{1,k+1}^n,U_{2,k+1}^n)\Big] \\
&= \sum_{k=1}^{n}\Big[H(Y_k|Y_1^{k-1}) - H(Y_k|W_1,U_{1,k+1}^n,Y_1^{k-1},W_2,U_{2,k+1}^n) - I(Y_k;U_{1,k+1}^n,U_{2,k+1}^n|W_1,W_2,Y_1^{k-1}) \\
&\qquad -H(U_{1k},U_{2k}|U_{1,k+1}^n,U_{2,k+1}^n) + H(U_{1k},U_{2k}|W_1,W_2,U_{1,k+1}^n,U_{2,k+1}^n)\Big] \\
&= \sum_{k=1}^{n}\Big[H(Y_k|Y_1^{k-1}) - H(Y_k|W_1,U_{1,k+1}^n,Y_1^{k-1},W_2,U_{2,k+1}^n) - I(U_{1k},U_{2k};Y_1^{k-1}|W_1,W_2,U_{1,k+1}^n,U_{2,k+1}^n) \\
&\qquad -H(U_{1k},U_{2k}) + H(U_{1k},U_{2k}|W_1,W_2,U_{1,k+1}^n,U_{2,k+1}^n)\Big] \\
&= \sum_{k=1}^{n}\Big[H(Y_k|Y_1^{k-1}) - H(Y_k|W_1,U_{1,k+1}^n,Y_1^{k-1},W_2,U_{2,k+1}^n) \\
&\qquad -H(U_{1k},U_{2k}) + H(U_{1k},U_{2k}|W_1,W_2,U_{1,k+1}^n,U_{2,k+1}^n,Y_1^{k-1})\Big] \\
&\leq \sum_{k=1}^{n}\Big[H(Y_k) - H(Y_k|W_1,U_{1,k+1}^n,W_2,U_{2,k+1}^n,Y_1^{k-1}) - I(U_{1k},U_{2k};L_{1k},L_{2k})\Big] \\
&= \sum_{k=1}^{n}\Big[I(Y_k;W_1,U_{1,k+1}^n,W_2,U_{2,k+1}^n,Y_1^{k-1}) - I(U_{1k},U_{2k};L_{1k},L_{2k})\Big] \\
&= \sum_{k=1}^{n}[I(Y_k;L_{1k},L_{2k}) - I(U_{1k},U_{2k};L_{1k},L_{2k})],
\end{aligned}
$$

which implies

$$R_1 + R_2 \leq \frac{1}{n}\sum_{k=1}^{n}[I(L_{1k},L_{2k};Y_k) - I(U_{1k},U_{2k};L_{1k},L_{2k})] + \epsilon_n. \tag{26}$$

We next introduce a time-sharing RV to simplify the bounds (23), (24), and (26) using a single-letter characterization. Define a RV V with alphabet $\{1,2,...,n\}$ and distribution $P_V(v) = 1/n$. We next introduce RVs U_1 and U_2 such that

$$\Pr(U_1 = u_1, U_2 = u_2) = \Pr(U_{1k} = u_1, U_{2k} = u_2) = Q_{U_1U_2}(u_1,u_2)$$

for all $(u_1,u_2) \in \mathcal{U}_1 \times \mathcal{U}_2$, which are independent of V. Furthermore, we define new RVs L_1, L_2, X_1, X_2, and Y by

$$
\begin{aligned}
&\Pr(L_1 = l_1, L_2 = l_2, X_1 = x_1, X_2 = x_2, Y = y|V = k) \\
&\quad = \Pr(L_{1k} = l_1, L_{2k} = l_2, X_{1k} = x_1, X_{2k} = x_2, Y_k = y)
\end{aligned}
$$

for all $(l_1, l_2, x_1, x_2, y) \in \mathcal{L}_1 \times \mathcal{L}_2 \times \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}$. It follows that

$$\begin{aligned}
&\frac{1}{n}\sum_{k=1}^{n}[I(L_{1k}; L_{2k}, Y_k) - I(U_{1k}; L_{1k})] \\
&= I(L_1; L_2, Y|V) - I(U_1; L_1|V) \\
&= H(L_1|V) - H(L_1|L_2, Y, V) - H(U_1|V) + H(U_1|L_1, V) \\
&\overset{(a)}{\leq} H(L_1) - H(L_1|L_2, Y, V) - H(U_1) + H(U_1|L_1, V) \\
&= I(L_1; L_2, Y, V) - I(U_1; L_1, V) \\
&\leq I(L_1, V; L_2, Y, V) - I(U_1; L_1, V) \\
&\overset{(b)}{=} I(T_1; T_2, Y) - I(T_1; U_1)
\end{aligned}$$

where (a) holds since conditioning reduces entropy and U_1 is independent of V, and in (b) $T_1 \triangleq (L_1, V)$ and $T_2 \triangleq (L_2, V)$. This shows that

$$R_1 \leq I(T_1; T_2, Y) - I(T_1; U_1) + \epsilon_n. \tag{27}$$

By a similar argument, we can show

$$R_2 \leq I(T_2; T_1, Y) - I(T_2; U_2) + \epsilon_n \tag{28}$$

and

$$R_1 + R_2 \leq I(T_1, T_2; Y) - I(U_1, U_2; T_1, T_2) + \epsilon_n. \tag{29}$$

For such RVs $(U_1, U_2, T_1, T_2, X_1, X_2, Y)$, it can be readily seen that the Markov chain relation $(U_1, U_2, T_1, T_2) \to (X_1, X_2) \to Y$ holds. In fact,

$$\begin{aligned}
&\Pr(Y = y|U_1 = u_1, U_2 = u_2, T_1 = t_1 = (l_1, k), T_2 = t_2 = (l_2, k), X_1 = x_1, X_2 = x_2) \\
&= \Pr(Y = y|U_1 = u_1, U_2 = u_2, L_1 = l_1, L_2 = l_2, X_1 = x_1, X_2 = x_2, V = k) \\
&= \Pr(Y_k = y|U_{1k} = u_1, U_{2k} = u_2, L_{1k} = l_1, L_{2k} = l_2, X_{1k} = x_1, X_{2k} = x_2) \\
&= \Pr(Y_k = y|X_{1k} = x_1, X_{2k} = x_2) \\
&= W_{Y|X_1X_2}(y|x_1, x_2).
\end{aligned}$$

Next we bound the distortions $\mathbb{E}[d_i(U_i, X_i)]$. Since (R_1, R_2) is achievable under the sequence of codes $(f_1^{(n)}, f_2^{(n)}, \psi^{(n)})$, this implies that for any $\delta > 0$ and all n large enough, we have

$$\begin{aligned}
D_i + \delta &\geq \frac{1}{n}\frac{1}{2^{nR_i}}\sum_{w_i=1}^{M_i}\sum_{\mathcal{U}_i^n} Q_{U_i}^{(n)}(\mathbf{u}_i) d_i\left(\mathbf{u}_i, f_i^{(n)}(w_i, \mathbf{u}_i)\right) \\
&= \frac{1}{n}\sum_{\mathcal{U}_i^n \times \mathcal{X}_i^n} \Pr(U_i^n = \mathbf{u}_i, X_i^n = \mathbf{x}_i) d_i(\mathbf{u}_i, \mathbf{x}_i) \\
&= \frac{1}{n}\sum_{k=1}^{n}\sum_{\mathcal{U}_i^n \times \mathcal{X}_i^n} \Pr(U_i^n = \mathbf{u}_i, X_i^n = \mathbf{x}_i) d_i(u_{ik}, x_{ik}) \\
&= \sum_{k=1}^{n} P_V(V = k) \sum_{\mathcal{U}_i \times \mathcal{X}_i} \Pr(U_{ik} = u_{ik}, X_{ik} = x_{ik}) d_i(u_{ik}, x_{ik})
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=1}^{n} P_V(V=k) \sum_{\mathcal{U}_i \times \mathcal{X}_i} \Pr(U_i = u_i, X_i = x_i | V = k) d_i(u_i, x_i) \\
&= \sum_{k=1}^{n} \sum_{\mathcal{U}_i \times \mathcal{X}_i} \Pr(U_i = u_i, X_i = x_i, V = k) d_i(u_i, x_i) \\
&= \sum_{\mathcal{U}_i \times \mathcal{X}_i} P_{U_i X_i}(u_i, x_i) d_i(u_i, x_i).
\end{aligned}
$$

Thus we obtained that $\mathbb{E}[d_i(U_i, X_i)] \leq D_i + \delta$ for $i = 1, 2$. Combined with (27)–(29) and recalling that $\lim_{n\to\infty} \epsilon_n = 0$ and that $\mathcal{R}(D_1, D_2)$ is closed, we conclude that $\mathcal{R}(D_1, D_2) \subset \mathcal{R}_{out}(D_1 + \delta, D_2 + \delta)$ as claimed. □

B Proof of Theorem 3

The forward part (achievability) is a consequence of Theorem 1 since (U_1, T_1) and (U_2, T_2) are independent and hence $I(T_1; T_2, Y) = I(T_1; Y|T_2)$, $I(T_2; T_1, Y) = I(T_2; Y|T_1)$, and $I(U_1, U_2; T_1, T_2) = I(U_1; T_1) + I(U_2; T_2)$. To prove the converse part, we need to sharpen the bounds in the last proof. We start from

$$
\begin{aligned}
& I(W_1; Y^n) - I(W_1; U_1^n) \\
&= \sum_{k=1}^{n} \Big[I(Y_k; W_1, U_{1,k+1}^n | Y_1^{k-1}) - I(U_{1k}; W_1, Y_1^{k-1}, U_{1,k+1}^n) \Big] \\
&= \sum_{k=1}^{n} \Big[H(W_1, U_{1,k+1}^n | Y_1^{k-1}) - H(W_1, U_{1,k+1}^n | Y_1^{k-1}, Y_k) - I(U_{1k}; W_1, Y_1^{k-1}, U_{1,k+1}^n) \Big] \\
&\overset{(a)}{=} \sum_{k=1}^{n} \Big[H(W_1, U_{1,k+1}^n | W_2, U_{2,k+1}^n, Y_1^{k-1}) - H(W_1, U_{1,k+1}^n | W_2, U_{2,k+1}^n, Y_1^{k-1}, Y_k) \\
&\qquad\qquad - I(U_{1k}; W_1, Y_1^{k-1}, U_{1,k+1}^n) \Big] \\
&= \sum_{k=1}^{n} \Big[I(W_1, U_{1,k+1}^n; Y_k | W_2, U_{2,k+1}^n, Y_1^{k-1}) - I(U_{1k}; W_1, Y_1^{k-1}, U_{1,k+1}^n) \Big] \\
&\leq \sum_{k=1}^{n} \Big[I(W_1, U_{1,k+1}^n, Y_1^{k-1}; Y_k | W_2, U_{2,k+1}^n, Y_1^{k-1}) - I(U_{1k}; W_1, Y_1^{k-1}, U_{1,k+1}^n) \Big] \\
&= \sum_{k=1}^{n} [I(L_{1k}; Y_k | L_{2k}) - I(U_{1k}; L_{1k})]
\end{aligned}
$$

where (a) follows since $(W_1, U_{1,k+1}^n)$ is now independent of $(W_2, U_{2,k+1}^n)$, and in the last equality we still let $L_{1k} \triangleq (W_1, Y_1^{k-1}, U_{1,k+1}^n)$ and $L_{2k} \triangleq (W_2, Y_1^{k-1}, U_{2,k+1}^n)$. Thus, using Fano's inequality we have

$$
R_1 \leq \frac{1}{n} \sum_{k=1}^{n} [I(L_{1k}; Y_k | L_{2k}) - I(U_{1k}; L_{1k})] + \epsilon_n.
$$

Similarly we can obtain

$$
R_2 \leq \frac{1}{n} \sum_{k=1}^{n} [I(L_{2k}; Y_k | L_{1k}) - I(U_{2k}; L_{2k})] + \epsilon_n.
$$

To bound the sum of the rates, we have

$$\begin{aligned} n(R_1+R_2) &= H(W_1,W_2) = I(W_1,W_2;Y^n) + H(W_1,W_2|Y^n) \\ &\leq I(W_1,W_2;Y^n) - I(W_1;U_1^n) - I(W_2;U_2^n) + n\epsilon_n \end{aligned} \tag{30}$$

and

$$\begin{aligned}
&I(W_1,W_2;Y^n) - I(W_1;U_1^n) - I(W_2;U_2^n) \\
&= \sum_{k=1}^{n} \Big[I(W_1;Y_k|Y_1^{k-1}) + I(W_2;Y_k|W_1,Y_1^{k-1}) - I(W_1;U_{1k}|U_{1,k+1}^n) - I(W_2;U_{2k}|U_{2,k+1}^n) \Big] \\
&= \sum_{k=1}^{n} \Big[H(Y_k|Y_1^{k-1}) - H(Y_k|W_1,Y_1^{k-1},U_{1,k+1}^n) - I(Y_k;U_{1,k+1}^n|W_1,Y_1^{k-1}) \\
&\qquad + H(Y_k|W_1,Y_1^{k-1}) - H(Y_k|W_1,W_2,Y_1^{k-1},U_{2,k+1}^n) - I(Y_k;U_{2,k+1}^n|W_1,W_2,Y_1^{k-1}) \\
&\qquad - H(U_{1k}|U_{1,k+1}^n) + H(U_{1k}|W_1,U_{1,k+1}^n) - H(U_{2k}|U_{2,k+1}^n) + H(U_{2k}|W_2,U_{2,k+1}^n) \Big] \\
&= \sum_{k=1}^{n} \Big[H(Y_k|Y_1^{k-1}) - H(Y_k|W_1,Y_1^{k-1},U_{1,k+1}^n) - I(U_{1k};Y_1^{k-1}|W_1,U_{1,k+1}^n) \\
&\qquad + H(Y_k|W_1,Y_1^{k-1}) - H(Y_k|W_1,W_2,Y_1^{k-1},U_{2,k+1}^n) - I(U_{2k};Y_1^{k-1}|W_1,W_2,U_{2,k+1}^n) \\
&\qquad - H(U_{1k}) + H(U_{1k}|W_1,U_{1,k+1}^n) - H(U_{2k}) + H(U_{2k}|W_1,W_2,U_{2,k+1}^n) \Big] \\
&= \sum_{k=1}^{n} \Big[H(Y_k|Y_1^{k-1}) - H(Y_k|W_1,Y_1^{k-1},U_{1,k+1}^n) \\
&\qquad + H(Y_k|W_1,Y_1^{k-1}) - H(Y_k|W_1,W_2,Y_1^{k-1},U_{2,k+1}^n) \\
&\qquad - H(U_{1k}) + H(U_{1k}|W_1,U_{1,k+1}^n,Y_1^{k-1}) - H(U_{2k}) + H(U_{2k}|W_1,W_2,U_{2,k+1}^n,Y_1^{k-1}) \Big] \\
&= \sum_{k=1}^{n} \Big[I(Y_k;W_1,U_{1,k+1}^n|Y_1^{k-1}) + I(Y_k;W_2,U_{2,k+1}^n|W_1,Y_1^{k-1}) \\
&\qquad - I(U_{1k};W_1,U_{1,k+1}^n,Y_1^{k-1}) - I(U_{2k};W_2,U_{2,k+1}^n,Y_1^{k-1}) \Big] \\
&= \sum_{k=1}^{n} \Big[H(W_1,U_{1,k+1}^n|Y_1^{k-1}) - H(W_1,U_{1,k+1}^n|Y_1^{k-1},Y_k) \\
&\qquad + H(W_2,U_{2,k+1}^n|W_1,Y_1^{k-1}) - H(W_2,U_{2,k+1}^n|W_1,Y_1^{k-1},Y_k) \\
&\qquad - I(U_{1k};W_1,U_{1,k+1}^n,Y_1^{k-1}) - I(U_{2k};W_2,U_{2,k+1}^n,Y_1^{k-1}) \Big] \\
&\overset{(a)}{=} \sum_{k=1}^{n} \Big[H(W_1,U_{1,k+1}^n|Y_1^{k-1}) - H(W_1,U_{1,k+1}^n|Y_1^{k-1},Y_k) \\
&\qquad + H(W_2,U_{2,k+1}^n|W_1,U_{1,k+1}^n,Y_1^{k-1}) - H(W_2,U_{2,k+1}^n|W_1,U_{1,k+1}^n,Y_1^{k-1},Y_k) \\
&\qquad - I(U_{1k};W_1,U_{1,k+1}^n,Y_1^{k-1}) - I(U_{2k};W_2,U_{2,k+1}^n,Y_1^{k-1}) \Big] \\
&= \sum_{k=1}^{n} \Big[H(W_1,U_{1,k+1}^n,W_2,U_{2,k+1}^n|Y_1^{k-1}) - H(W_1,U_{1,k+1}^n,W_2,U_{2,k+1}^n|Y_1^{k-1},Y_k) \\
&\qquad - I(U_{1k};W_1,U_{1,k+1}^n,Y_1^{k-1}) - I(U_{2k};W_2,U_{2,k+1}^n,Y_1^{k-1}) \Big]
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=1}^{n} \Big[I(W_1, U_{1,k+1}^n, W_2, U_{2,k+1}^n; Y_k | Y_1^{k-1}) - I(U_{1k}; W_1, U_{1,k+1}^n, Y_1^{k-1}) - I(U_{2k}; W_2, U_{2,k+1}^n, Y_1^{k-1}) \Big] \\
&\leq \sum_{k=1}^{n} \Big[H(Y_k) - H(Y_k | W_1, U_{1,k+1}^n, W_2, U_{2,k+1}^n, Y_1^{k-1}) - I(U_{1k}; W_1, U_{1,k+1}^n, Y_1^{k-1}) \\
&\qquad\qquad - I(U_{2k}; W_2, U_{2,k+1}^n, Y_1^{k-1}) \Big] \\
&= \sum_{k=1}^{n} \Big[I(W_1, U_{1,k+1}^n, W_2, U_{2,k+1}^n, Y_1^{k-1}; Y_k) - I(U_{1k}; W_1, U_{1,k+1}^n, Y_1^{k-1}) - I(U_{2k}; W_2, U_{2,k+1}^n, Y_1^{k-1}) \Big] \\
&= \sum_{k=1}^{n} [I(L_{1k}, L_{2k}; Y_k) - I(U_{1k}; L_{1k}) - I(U_{2k}; L_{2k})]
\end{aligned}
$$

where (a) holds since $(W_1, U_{1,k+1}^n)$ is independent of $(W_2, U_{2,k+1}^n)$ and $L_{1k} \triangleq (W_1, Y_1^{k-1}, U_{1,k+1}^n)$ and $L_{2k} \triangleq (W_2, Y_1^{k-1}, U_{2,k+1}^n)$ in the last equality. The above implies

$$
R_1 + R_2 \leq \frac{1}{n} \sum_{k=1}^{n} [I(L_{1k}, L_{2k}; Y_k) - I(U_{1k}; L_{1k}) - I(U_{2k}; L_{2k})] + \epsilon_n.
$$

The rest of the proof proceeds the same way as the proof of Theorem 2. □

C Upper Bounds on $|\mathcal{T}_i|$ for $\mathcal{R}_{in}^*(D_1, D_2)$ and $\mathcal{R}_{in}(D_1, D_2)$

We only bound the cardinality of $\mathcal{T}_1$ and $\mathcal{T}_2$ for the region $\mathcal{R}_{in}^*(D_1, D_2)$. The bounds for $|\mathcal{T}_1|$ and $|\mathcal{T}_2|$ for the region $\mathcal{R}_{in}(D_1, D_2)$ can be derived in a similar manner. We will need the following support lemma, which is based on Carathéodory's theorem on the convex hull of a set in a finite-dimensional vector space.

Lemma 4 ([2, Support lemma, p. 311]) Let $\mathcal{P}(\mathcal{X})$ be the set of distributions defined on a finite set $\mathcal{X}$ (represented as the probability simplex in $\mathbb{R}^{|\mathcal{X}|}$) and let f_j, $j = 1, 2, ..., k$ be real-valued continuous functions on $\mathcal{P}(\mathcal{X})$. For any probability measure μ on the Borel σ-algebra of $\mathcal{P}(\mathcal{X})$, there exist k elements $P_1, P_2, ..., P_k$ of $\mathcal{P}(\mathcal{X})$ and k non-negative reals $\alpha_1, \alpha_2, ...\alpha_k$ with $\sum_{i=1}^{k} \alpha_i = 1$ such that for every $j = 1, 2, ..., k$

$$
\int_{\mathcal{P}(\mathcal{X})} f_j(P) \mu(dP) = \sum_{i=1}^{k} \alpha_i f_j(P_i).
$$

Using this lemma, we will show that for any given $P_{X_1 T_1 | U_1}$ and $P_{X_2 T_2 | U_2}$, there exists a RV $\widehat{T}_1$ with $|\widehat{\mathcal{T}}_1| \leq |\mathcal{U}_1||\mathcal{X}_1| + 1$ only depending on U_1 and X_1 such that the following hold

$$
\begin{aligned}
I(\widehat{T}_1; Y | T_2) - I(U_1; \widehat{T}_1) &= I(T_1; Y | T_2) - I(U_1; T_1) && (31) \\
I(T_2; Y | \widehat{T}_1) - I(U_2; T_2) &= I(T_2; Y | T_1) - I(U_2; T_2) && (32) \\
I(\widehat{T}_1, T_2; Y) - I(U_1; \widehat{T}_1) - I(U_2; T_2) &= I(T_1, T_2; Y) - I(U_1; T_1) - I(U_2; T_2), && (33)
\end{aligned}
$$

and that the expectation of the distortion between U_1 and X_1 is preserved when T_1 is replaced by $\widehat{T}_1$. Note that the upper bound on $|\widehat{\mathcal{T}}_1|$ does not depend on $|\mathcal{T}_2|$.

We first rewrite

$$\begin{aligned} I(T_1;Y|T_2) - I(U_1;T_1) &= H(Y|T_2) - H(Y|T_1,T_2) - H(U_1) + H(U_1|T_1), \\ I(T_2;Y|T_1) - I(U_2;T_2) &= H(Y|T_1) - H(Y|T_1,T_2) - I(U_2;T_2), \end{aligned}$$

and

$$I(T_1,T_2;Y) - I(U_1;T_1) - I(U_2;T_2) = H(Y) - H(Y|T_1,T_2) - H(U_1) + H(U_1|T_1) - I(U_2;T_2).$$

Recall that the joint distribution of $(U_1,U_2,T_2,T_2,X_1,X_2,Y)$ can be factorized as

$$P_{U_1T_1U_2T_2X_1X_2Y} = Q_{U_1U_2}P_{T_1X_1|U_1}P_{T_2X_2|U_2}W_{Y|X_1X_2}.$$

We note that there exists a Markov chain $(T_1,X_1) \to U_1 \to U_2 \to (T_2,X_2)$. Writing

$$P_{U_1T_1U_2T_2X_1X_2Y} = P_{T_1}P_{U_1X_1|T_1}P_{U_2|U_1}P_{T_2X_2|U_2}W_{Y|X_1X_2},$$

and noting that $P_{U_2|U_1}$, $P_{T_2X_2|U_2}$ and $W_{Y|X_1X_2}$ are fixed, to apply the support lemma, we need $m-1$ functions to preserve the joint distribution of (U_1,X_1) (see (34) below), where $m \triangleq |\mathcal{U}_1||\mathcal{X}_1|$. Specifically, we define the following real-valued continuous functions of distribution $P_{U_1X_1|T_1}(\cdot,\cdot|t_1)$ on $\mathcal{U}_1 \times \mathcal{X}_1$ for fixed $t_1 \in \mathcal{T}_1$,

$$f_{u_1,x_1}(P_{U_1X_1|T_1}(\cdot,\cdot|t_1)) \triangleq P_{U_1X_1|T_1}(u_1,x_1|t_1)$$

for all $(u_1,x_1) \in \mathcal{U}_1 \times \mathcal{X}_1$ except one pair (u_1,x_1). Furthermore, we define real-valued continuous functions

$$\begin{aligned} f_m(P_{U_1X_1|T_1}(\cdot,\cdot|t_1)) &\triangleq -H_P(Y|T_1=t_1,T_2) + H_P(U_1|T_1=t_1), \\ f_{m+1}(P_{U_1X_1|T_1}(\cdot,\cdot|t_1)) &\triangleq H_P(Y|T_1=t_1) - H_P(Y|T_1=t_1,T_2), \end{aligned}$$

where the entropies are taken under the joint distribution induced by $P_{U_1X_1|T_1}(\cdot,\cdot|t_1)$. According to the support lemma, there must exist a new RV $\widehat{T}_1$ (jointly distributed with (U_1,X_1)) with alphabet size $|\widehat{T}_1| = m+1 = |\mathcal{U}_1||\mathcal{X}_1|+1$ such that the expectation of f_i, $i=1,2,...,m+1$, with respect to P_{T_1} can be expressed in terms of the convex combination of $m+1$ points; i.e.,

$$\begin{aligned} P_{U_1X_1}(u_1,x_1) &= \sum_{t_1 \in \mathcal{T}_1} P_{T_1}(t_1) f_{u_1,x_1}(P_{U_1X_1|T_1}(\cdot,\cdot|t_1)) \\ &= \sum_{\widehat{t}_1 \in \widehat{\mathcal{T}}_1} P_{\widehat{T}_1}(\widehat{t}_1) f_{u_1,x_1}(P_{U_1X_1|\widehat{T}_1}(\cdot,\cdot|\widehat{t}_1)), \end{aligned} \tag{34}$$

$$\begin{aligned} -H(Y|T_1,T_2) + H(U_1|T_1) &= \sum_{t_1 \in \mathcal{T}_1} P_{T_1}(t_1) f_m(P_{U_1X_1|T_1}(\cdot,\cdot|t_1)) \\ &= \sum_{\widehat{t}_1 \in \widehat{\mathcal{T}}_1} P_{\widehat{T}_1}(\widehat{t}_1) f_m\left(P_{U_1X_1|\widehat{T}_1}(\cdot,\cdot|\widehat{t}_1)\right) \\ &= -H(Y|\widehat{T}_1,T_2) + H(U_1|\widehat{T}_1), \end{aligned}$$

$$\begin{aligned} H(Y|T_1) - H(Y|T_1,T_2) &= \sum_{t_1 \in \mathcal{T}_1} P_{T_1}(t_1) f_{m+1}(P_{U_1X_1|T_1}(\cdot,\cdot|t_1)) \\ &= \sum_{\widehat{t}_1 \in \widehat{\mathcal{T}}_1} P_{\widehat{T}_1}(\widehat{t}_1) f_{m+1}(P_{U_1X_1|\widehat{T}_1}(\cdot,\cdot|\widehat{t}_1)) \\ &= H(Y|\widehat{T}_1) - H(Y|\widehat{T}_1,T_2). \end{aligned}$$

This implies that (31)–(33) hold. It should be point out that this RV $\widehat{T}_1$ maintains the prescribed distortion level, since $P_{U_1X_1}(u_1, x_1)$ is preserved. Similarly, for any given $P_{X_1T_1|U_1}$ and $P_{X_2T_2|U_2}$, we can show that there exists a RV $\widehat{T}_2$ with $|\widehat{\mathcal{T}}_2| \leq |\mathcal{U}_2||\mathcal{X}_2| + 1$ only depending on U_2 and X_2 such that

$$\begin{aligned} I(T_1; Y|\widehat{T}_2) - I(U_1; T_1) &= I(T_1; Y|T_2) - I(U_1; T_1) && (35) \\ I(\widehat{T}_2; Y|T_1) - I(U_2; \widehat{T}_2) &= I(T_2; Y|T_1) - I(U_2; T_2) && (36) \\ I(T_1, \widehat{T}_2; Y) - I(U_1; T_1) - I(U_2; \widehat{T}_2) &= I(T_1, T_2; Y) - I(U_1; T_1) - I(U_2; T_2), && (37) \end{aligned}$$

and the distortion constraint between U_2 and X_2 is preserved. Thus we conclude that the cardinality of $\mathcal{T}_i$ can be bounded by $|\mathcal{U}_i||\mathcal{X}_i| + 1$, $i = 1, 2$.

Finally, we remark that the support lemma cannot be straightforwardly used to bound the cardinality for $\mathcal{T}_1$ and $\mathcal{T}_2$ for the region $\mathcal{R}_{out}(D_1, D_2)$ and $\mathcal{R}^*_{out}(D_1, D_2)$. For example, to bound the cardinality of $\mathcal{T}_1$ for $\mathcal{R}_{out}(D_1, D_2)$, we need $|\mathcal{U}_1||\mathcal{U}_2||\mathcal{X}_1||\mathcal{X}_2||\mathcal{T}_2| - 1$ real-valued continuous functions to preserve the joint distribution of $(U_1, U_2, T_2, X_1, X_2)$. Therefore, we may need $|\mathcal{U}_1||\mathcal{U}_2||\mathcal{X}_1||\mathcal{X}_2||\mathcal{T}_2| + 1$ letters and this upper bound depends on $|\mathcal{T}_2|$. □

References

[1] A. S. Cohen and A. Lapidoth, "The Gaussian watermarking game," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1639–1667, Jun. 2002.

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

[3] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inform. Theory*, vol. 24, pp. 339–348, May 1978.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2^{nd} Edition, Wiley, 2006.

[5] S. Gelfand and M. Pinsker, "Coding for a channel with random parameters," *Problems of Control and Information Theory*, vol. 9, pp. 19–31, 1980.

[6] T. S. Han and K. Kobayashi, "A unified achievable rate region for a general class of multiterminal source coding systems," *IEEE Trans. Inform. Theory*, vol. 26, no. 3, pp. 396–412, May 1980.

[7] S. Kotagiri and J. N. Laneman, "Reversible information embedding in multi-user channels," *Proc. Allerton Conf. Communications, Control, and Computing*, Monticello, IL, Sept. 2005.

[8] S. Kotagiri and J. N. Laneman, "Variations on information embedding in multiple access and broadcast channels," *IEEE Trans. Inform. Theory*, to appear. Available at http://www.nd.edu/ jnl/pubs/it2007c.pdf

[9] P. Moulin and J. O'Sullivan, "Information-theoretic analysis of information hiding," *IEEE Trans. Inform. Theory*, vol. 49, no. 3, pp. 563–593, Mar. 2003.

[10] P. Moulin and M. K. Mihcak, " The parallel-Gaussian watermarking game," *IEEE Trans. Inform. Theory*, vol. 50, no. 2, pp. 272–289, Feb. 2004.

[11] Y. Oohama, "Gaussian multiterminal source coding," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1912–1923, Nov. 1997.

[12] W. Sun and E. H. Yang, "On achievable regions of public multiple-access Gaussian watermarking systems," *Proc. 6th Int. Inform. Hiding Workshop*, Toronto, Canada, May 23–25, 2004.

[13] S. Y. Tung, "Multiterminal source coding," Ph.D dissertation, School of Electrical Engineering, Cornell Univ., Ithaca, NY, May 1978.

[14] Yadong Wang, *Hybrid Digital-Analog Source-Channel Coding and Information Hiding: Information-Theoretic Perspectives*, Ph.D. thesis, Dept. of Math. and Stat., Queen's University, Kingston, ON, Canada, Sept. 2007.

[15] F. M.J. Willems, "An information-theoretical approach to information embedding," Proc. *21st Symposium on Information Theory*, pp. 255–260, Wassenaar, The Netherlands, May 25–26, 2000.