Plausible Deniability over Broadcast Channels

Mayank Bakshi* Vinod Prabhakaran†

Abstract

In this paper, we introduce the notion of Plausible Deniability in an information theoretic framework. We consider a scenario where an entity that eavesdrops through a broadcast channel summons one of the parties in a communication protocol to reveal their message (or signal vector). It is desirable that the summoned party have enough freedom to produce a fake output that is likely plausible given the eavesdropper’s observation. We examine three variants of this problem – Message Deniability, Transmitter Deniability, and Receiver Deniability. In the first setting, the message sender is summoned to produce the sent message. Similarly, in the second and third settings, the transmitter and the receiver are required to produce the transmitted codeword, and the received vector respectively. For each of these settings, we examine the maximum communication rate that allows a given minimum rate of plausible fake outputs. For the Message and Transmitter Deniability problems, we fully characterise the capacity region for general broadcast channels, while for the Receiver Deniability problem, we give an achievable rate region for stochastically degraded broadcast channels.

I. Introduction

Communicating reliably and securely is increasingly being viewed as a challenging task in a world where the connected nature of communication networks make eavesdropping easier than ever before. Further, information leakage during communication may have a wide range of consequences depending on the usage scenario – for a patient transmitting their medical data, the leakage may amount to loss of privacy, while for a whistleblower, this may have dire (even life-threatening) consequences. These concerns have given rise to a wide variety of secure communication protocols – two notable paradigms being the cryptographic approach [1] and the information-theoretic approach [2, 3]. In a typical such system, the sender encrypts the message using a previously agreed upon protocol and the receiver decrypts it accordingly, while simultaneously guaranteeing that the maximum amount of potential information that would be leaked to the eavesdropper is bounded by a security or equivocation parameter.

We argue that in the above protocols, it is often assumed the eavesdropper may passively or actively eavesdrop on the transcript, but never communicates directly with the sender or the receiver. However, as often seen in the case of the whistleblower, in addition to observing the communication, the eavesdropping entity can also actively interact with the sender or receiver and force them to divulge either the message or the codeword. In this situation, it is desirable that the whistleblower be able to protect his or her right to free speech and produce a fake message (or codeword) with as little dependence on the true message (or codeword) as possible, while also appearing to be truthful with respect to the eavesdropper’s observation. The following example highlights how this requirement differs from the usual notion of secrecy.

Example (Secrecy does not guarantee freedom of expression). Consider the setting of Figure [1] Since the channel to Bob is noiseless, the secrecy capacity [3] is $p$. On the other hand, even if Alice and Bob operate a code equipped with an information-theoretic secrecy guarantee and Judy demands that Alice provide the transmitted codeword $x$, Alice has no choice but to provide exactly what was transmitted (and hence, also reveal the message). If Alice chooses to provide a vector $x'$ different from $x$, then Judy would be able to detect with a constant probability that Alice is lying since the transmitted symbol for any coordinate where $x'$ and $x$ differ would be received correctly by Judy with probability $1 - p$.

The above example shows that secrecy with respect to eavesdropping may not be sufficient when one of the communicating parties is summoned to reveal their observations. Recognising this limitation, the notion of Plausible Deniability [4] has recently garnered much attention in the cryptographic community. Generally speaking, communication schemes based on this principle allow the summoned party to pretend deniability by responding with a fake message while simultaneously appearing to be plausible with respect to any eavesdropping. By now there are both good algorithms [5, 6] as well as practical implementations [7] of this principle that rely on computational assumptions on the eavesdropper.

In this work, we examine Plausible Deniability in an information theoretic setting and allow all parties to be computationally unbounded. Our general setup is as follows. Alice, Bob, and Charlie are three participants in a potentially secretive communication setup. Charlie wishes to send a message $m \in \mathcal{M}$ to Bob through Alice. Alice and Bob are at two ends of a noisy channel and operate the physical layer with Alice being the transmitter and Bob being the receiver, while Charlie interacts directly with Alice and knows the message but does not partake in the physical layer transmission and reception. The nature of the message may either be an innocuous or a secretive one – this is known to Alice, Bob, and Charlie, but not to any eavesdroppers.

---

*Mayank Bakshi (email: mayank@inc.cuhk.edu.hk) is affiliated with the Institute of Network Coding, The Chinese University of Hong Kong, Hong Kong. The work described in this paper was partially supported by a grant from University Grants Committee of the Hong Kong Special Administrative Region, China (Project No. AoE/08/08).

†Vinod Prabhakaran (email: vinodmp@tif.rsl.in) is affiliated with the Tata Institute of Fundamental Research, India. Vinod Prabhakaran’s research was funded in part by a Ramnajan fellowship from the Department of Science and Technology, Government of India and in part by Information Technology Research Academy (ITRA), Government of India under ITRA-Mobile grant ITRA/15(64)/Mobile/USEAADWN/01.
Fig. 1: Alice wishes to communicate a message $m$ to Bob by sending a codeword $x$ over a noiseless binary channel while an eavesdropper Judy observes $x$ through a binary erasure channel with erasure probability $p > 0$. Note that, in order to avoid being detected as lying, the summoned party’s output should appear plausible to Judy given her side information $z$. In particular, for the channel in this example, both Alice and Bob are forced to reveal their true codewords (i.e., $x$) to Judy. This example also shows a contrast between the standard notion of secrecy and the plausible deniability requirement.

Judy is an eavesdropper who observes a noisy version of Alice’s transmission. In this work, we assume that the statistics of Judy’s observation are known to the above three parties, but the exact observation is unknown. We consider three settings for this problem. In the Transmitter Deniability problem, Judy may summon Alice and ask her to produce the transmitted codeword. Similarly, in the Receiver Deniability, and the Message Deniability problems, Judy may summon Bob, and Charlie, to produce the received vector, and the message, respectively. In each of these settings, depending on whether the communication is innocuous or secretive, the summoned party may either respond truthfully or use a Faking Procedure to produce a fake output that reveals as little information about the true message as possible while still maintaining plausibility with respect to Judy’s observation.

We quantify the efficacy of a communication scheme in terms of its two properties – the reliability of the code and the plausible deniability of the faking procedure. The first property i.e., the reliability is measured in a standard fashion in terms of the message rate and the error probability at the decoder. Plausible deniability is also measured in terms of two metrics – the plausibility and the rate of deniability. Roughly speaking, plausibility measures the closeness between two distributions – the joint distribution of the fake output with the eavesdropper’s observation and that of the true message or signal vector with the eavesdropper’s observation. We measure this distance in terms of the Kullback-Leibler (K-L) divergence.\(^1,2\) The rate of deniability is measured as the conditional entropy of the fake message given the summoned party’s observations. This attempts to capture the amount of freedom the summoned party has while responding to the summons. The rate of deniability may also be roughly interpreted as a measure of equivocation at the eavesdropper after the summoned party is forced to respond. Strictly speaking, the rate of deniability is a purely operational characteristic of the faking procedure and our formal definition of the rate of deniability does not appear to be related to equivocation. However, when the faking procedure satisfies the plausibility requirement, we establish an asymptotic equivalence between these two notions in Propositions\(^2\) and \(^3\). We also emphasize here that demanding a rate of deniability $D$ is a stronger requirement than demanding an equivocation $D$ in the usual information theoretic secrecy setting – this naturally extends similar observations in the cryptographic setting where, a plausibly deniable protocol trivially also satisfies the security requirement. The rest of this paper is organised as follows. In Section II we formally describe our notation and problem formulation and state the main results in Section III. In Sections IV and V we give proof sketches for our theorems, and discuss some examples and key properties of our capacity regions. Finally, in Section VI we provide concluding remarks.
II. Problem Formulation

a) Notation: Throughout this paper, we typically adopt the following notation. Upper case math and lower case symbols such as $X$ and $x$ denote random variables and their specific values respectively. Boldface symbols such as $\mathbf{X}$ and $\mathbf{x}$ denote random vectors and their specific values respectively, while calligraphic symbols such as $\mathcal{X}$ denote sets. Probability distributions of generic random variables is typically written as $P$ (e.g. $P_X$, $P_{Y|X}$), while probability distributions imposed by the specific codebook are typically written as $Q$ (e.g. $Q_X$). Throughout this paper, we employ strong typicality in our analysis, and define the strongly typical set for a random variable $X$ as

$$A^n_T(X) = \left\{ x \in \mathcal{X}^n : \max_{x \in \mathcal{X}} \left[ \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{x_i = x\} - P_X(x) \right] \leq \epsilon \right\}. $$

b) Channel model: Alice, Bob, and Judy are connected through the following memoryless broadcast channel – at any time instant, Alice’s transmission $X$ is used to convey Bob’s reception $Y$ and Judy’s observation $Z$.

In the hypothesis testing sense.

$$\delta \equiv \{ (m,k_a,k_b,k_c) : m \in \mathcal{M}, k_a \in \mathcal{K}_A, k_b \in \mathcal{K}_B, k_c \in \mathcal{K}_C \}.$$ Note that $\mathcal{C}$ is a multi-set with possible repetitions as we do not require that $\text{Enc} \cdot \mathcal{M}$ to be injective. We say that $\mathcal{C}$ is $(\epsilon,R)$-reliable if $\frac{1}{n} \log_2 |\mathcal{M}| \geq R$, and there exists an encoder and decoder pair $(\text{Enc},\text{Dec})$ such that $Q_{\mathcal{M}XY}(\mathcal{M} \neq \mathcal{M}) < \epsilon$.

c) Codes: A code is a pair of maps $(\text{Enc},\text{Dec})$ which are applied by Alice and Bob to generate the codeword $x \in X^n = \text{Enc}(m,k_a)$ and the reconstruction $\hat{m} = \text{Dec}(y)$ respectively. When there is no private randomness at Alice, we denote the codeword for message $m$ by $x(m)$. To simplify notation, we represent a code $(\text{Enc},\text{Dec})$ through its codebook $\mathcal{C} \equiv \{ (m,k_a,k_b,k_c) : m \in \mathcal{M}, k_a \in \mathcal{K}_A, k_b \in \mathcal{K}_B, k_c \in \mathcal{K}_C \}$. In addition, the code and the faking procedure (defined in the following) are known to all parties.

d) Plausible deniability: Judy may summon Alice, Bob, or Charlie to provide a variable $w \in \mathcal{W}$ that can be used to reconstruct the message using a map $\text{Msg} : \mathcal{W} \rightarrow \mathcal{M}$. Depending on whether or not the transmission is an innocuous, the summoned party may either reveal the true value of $w$ or use a faking procedure $\text{Fake}(\cdot)$ that accepts as input the true value $w$ along with any private randomness, to output a fake value $w^\dagger \in \mathcal{W}$. We characterize a faking procedure through its plausibility $\delta$ and the deniability rate $D$. In general, we say that $w^\dagger$ is $(\delta,D)$-plausibly deniable for $w$ given observation $Z$ if

$$(i) \quad D(Q_{\mathcal{W}|w^\dagger} || Q_{\mathcal{W}|w}) \leq \delta, \quad \text{and} \quad (ii) \quad \frac{1}{n} H(\text{Msg}(w^\dagger) | w) \geq D.$$ In the various settings considered in this paper, $w$ equals $x,y$, or $m$. The three settings we consider in this paper are:

(A) Message deniability: As shown in Figure 2 Charlie is the summoned party, $w = m$, $\mathcal{W} = \mathcal{M}$, and $\text{Msg}(w) = w$.

(B) Transmitter deniability: Here, Alice is the summoned party, $w = x$, $\mathcal{W} = \mathcal{X}^n$, and $\text{Msg}(w)$ is the most likely guess given that $x = w$, i.e., $\text{Msg}(w) \equiv \arg\max_{m \in \mathcal{M}} Q_{\mathcal{M}X}(m|w)$ if the maximum is attained at a unique value of $m$. If there are multiple values of $m$ achieving the above maximum, then $\text{Msg}(w)$ selects one of them arbitrarily.

(C) Receiver deniability: Here, Bob is the summoned party, $w = y$, $\mathcal{W} = \mathcal{Y}^n$, and $\text{Msg}(w) = \text{Dec}(w)$.

e) Capacity regions: For each setting $w \in \{m,x,y\}$, we say that a rate-deniability pair $(R,D)$ is achievable if for any $\epsilon,\delta > 0$ and large enough $n$, there exists a blocklength-$n$ code $\mathcal{C}$ that is $(\epsilon,R)$-reliable and a faking procedure $\text{Fake}(\cdot)$ that is $(\delta,D)$-plausibly deniable for $w$ given $Z$. The capacity region $\mathcal{R}_w$ is the closure of the set of all achievable rate-deniability pairs.

III. MAIN RESULTS

Theorem 1 (Message Deniability). $\mathcal{R}_m$ is the set of all $(R,D)$ pairs such that there exist random variables $U \in \mathcal{Y}$ and $V \in \mathcal{Y}$ satisfying $V - U - X - (Y,Z)$, $\mathcal{Y} \notin \mathcal{X}$, and

$$0 \leq R \leq I(Y;V) + I(U;Y|V) - I(U;Z|V),$$

$$0 \leq D \leq I(U;Y|V) - I(U;Z|V).$$

Theorem 2 (Transmitter Deniability). $\mathcal{R}_x$ is the set of all $(R,D)$ pairs such that there exists a random variable $U$ satisfying $U - X - (Y,Z)$ and $X - U - Z$, and

$$0 \leq R \leq I(X;Y), \quad \text{and} \quad 0 \leq D \leq I(X;Y).$$

Theorem 3 (Achievability for Receiver Deniability). Let $P_{XY|Z}$ be a stochastically degraded broadcast channel, i.e., $P_{XY|Z}(y|x) = \sum_{z \in \mathcal{Y}} P_{Z|Y}(z|y)P_{X|Y}(x|z)$ for some distribution $P_{Z|Y}$. Then, $\mathcal{R}_y$ includes all $(R,D)$ pairs such that there exists a random variable $V$
satisfying $V - Y - (X, Z)$, and $Y - V - Z$, and
\[ 0 \leq R \leq I(X; Y), \text{ and } 0 \leq D \leq I(X; Y|V). \]

IV. MESSAGE DENIABILITY

In this section, we outline the proof of Theorem 1. Due to the standard nature of the achievability proof, we only give a proof sketch here, and instead focus mainly on the converse. Our achievability argument relies on reducing our problem to the following variant of the information theoretic secrecy problem.

Secrecy with side information at the eavesdropper: Consider the setup shown in Figure 3. Alice observes independent sources $s \in \{0,1\}^{nR_s}$ and $t \in \{0,1\}^{nR_t}$ and wishes to transmit them reliably to Bob over $n$ uses of the channel. Judy observes a noisy version of the transmission and knows the source $t$ as side information. The goal for the transmission is to ensure that the leakage $I(S;Z|T)$ is small. At first sight, the setting here is similar to the public message and confidential message setting of [3]. However, in contrast to [3], Judy is not interested in estimating $t$ based on $z$, but is instead provided with $t$ as side-information. This allows us to operate at potentially higher rates than [3]. By using a construction based on superposition coding and following similar arguments as [3], all rate pairs $(R_s, R_t)$ such that $R_s \leq I(U; Y|V) - I(U; Z|V)$, and $R_t \leq I(V; Y)$ for some $(U, V)$ satisfying $V - U - X - (Y, Z)$ are achievable. The full proof of the above is skipped here.

A. Proof sketch of achievability of Theorem 1

The crux of the achievability proof is the following reduction argument. For the message deniability problem, we decompose the $nR$-length message $m$ into two parts – a confidential part $m_c$ of $nR_s$ bits, and a public part $m_p$ of $nR_t$ bits. Next, Alice and Bob encode and decode $(m_c, m_p)$ using a reliable code $(\text{Enc}_c, \text{Dec}_c)$ for the above mentioned secrecy problem. The reliability guarantees for our code thus follow directly. The faking procedure draws $M'_c$ uniformly at random $\{0,1\}^{nR_s}$ and outputs $(m'_c, m_p)$. Based on the information leakage guarantees of the code $(\text{Enc}_c, \text{Dec}_c)$, we conclude that the plausible deniability requirement is satisfied with $R = R_s + R_t$ and $D = R_s$. This shows that $(R_s + R_t, R_s) \in \sigma_m$, and leads to the achievability of all rates claimed in Theorem 1.

B. Proof of converse in Theorem 1

Before stating the formal proof, we state the following lemma.

Lemma 1. Let $C$ be $(\epsilon, R)$-reliable code and let $M''$ be $(\delta, D)$-plausibly deniable for $M$. Then, there exists non-negative constants $\lambda$ depending only on $P_{ZX}$ and $R$ such that
\[ I(M; Z|M'') \leq \delta + n\lambda \sqrt{\delta}, \text{ and } \]
\[ I(H(M) - H(M'')) \leq \delta + n\lambda \sqrt{\delta}. \]

Proof: We explicitly only prove the first inequality. The second inequality follow from a similar reasoning.

\[ I(M; Z|M'') = H(Z|M'') - H(Z|M) \]
\[
\begin{align*}
&= \sum_{(z, m) : q_{M^0}(z, m) > 0} Q_{\overline{M}^0}(z, m) \log_2 \frac{Q_{\overline{M}^0}(z, m)}{Q_{\overline{M}}(z, m)} - \sum_{m : \overline{M}^0(m) > 0} Q_{\overline{M}}(m) \log_2 \frac{Q_{\overline{M}}(m)}{Q_{\overline{M}}(z, m)} \\
&= \sum_{m : Q_{\overline{M}}(m) > 0} Q_{\overline{M}}(m) \log_2 Q_{\overline{M}}(m) - \sum_{m : Q_{\overline{M}}(m) > 0} Q_{\overline{M}}(m) \log_2 Q_{\overline{M}}(m)
\end{align*}
\]

\[
\leq \delta - \sum_{x : Q_{\overline{M}}(x) > 0} \left[ Q_{\overline{M}^0}(z, m) - Q_{\overline{M}}(z, m) \right] \log_2 \frac{1}{Q_{X, M}^{(z, m)}}
\]

\[
\leq \delta + \sum_{x : Q_{\overline{M}}(x) > 0} \left[ Q_{\overline{M}^0}(z, m) - Q_{\overline{M}}(z, m) \right] \log_2 \frac{1}{Q_{X, M}^{(z, m)}}
\]

\[
\leq \delta + \sum_{x : Q_{\overline{M}}(x) > 0} \left[ Q_{\overline{M}^0}(z, m) - Q_{\overline{M}}(z, m) \right] \sum_{x : Z,M^0(x) > 0} \log_2 \frac{1}{P_{X}(x)}
\]

\[
\leq \delta + \sum_{x : Q_{\overline{M}}(x) > 0} \left[ Q_{\overline{M}^0}(z, m) - Q_{\overline{M}}(z, m) \right] \sum_{x : Z,M^0(x) > 0} \log_2 \frac{1}{P_{X}(x)}
\]

\[
\leq \delta + n \sqrt{\delta} \left[ R - \log_2 \frac{1}{\min_{x : P_{X}(x) > 0}} \right]
\]

In the above, \((a)\) follows by using the fact that \(M^0\) is \((\delta, D)\)-plausible deniable for \(M\) given \(Z\) to bound the first term in \([1]\), noting that \(P_{M}(m)\) equals \(1/|\mathcal{M}|\) to conclude that the second term is zero, and applying the non-negativity of the Kullback-Leibler divergence. The inequality \((b)\) is obtained by using Jensen's inequality. Finally, \((c)\) follows applying Pinsker's inequality to bound the variational distance between the distributions \(Q_{\overline{M}^0}^{(z, m)}\) and \(Q_{\overline{M}^{(z, m)}}\).

**Proof of converse of Theorem \([7]\).** We begin by obtaining \(n\)-letter bounds on \(D\) and \(R\) for any \((\epsilon, R)\)-reliable and \((\delta, D)\)-plausibly deniable code. To this end, from the definition and Lemma \([1]\), there exists \(\gamma = \gamma(\epsilon, \delta) > 0\) such that \(\lim_{(\epsilon, \delta) \to (0, 0)} \gamma = 0\), and

\[
nD = H(M^0|\overline{M})
= H(M^0) + I(M; Y | M^0) - H(M)
\leq H(M^0) + n\gamma
\leq I(M; Y | M^0) + 2n\gamma
\leq I(M; Y | M^0) - I(M; Z | M^0) + 3n\gamma.
\]

Next, from Fano's inequality,

\[
nR \leq I(M; Y) + n\gamma
= I(M^0, M; Y) + n\gamma
= I(M^0; Y) + I(M; Y | M^0) + n\gamma
\leq I(M^0; Y) + I(M; Y | M^0) - I(M; Z | M^0) + 2n\gamma,
\]

where the second equality follows from the fact that \(M^0 - M - Y\) is a Markov chain. Next, we obtain single-letter versions of the above expressions. Let \(T\) be uniformly distributed over \([1 : n]\) and independent of \((M, M^0, X, Y, Z)\). From \([2]\),

\[
D \leq \frac{1}{n} \left[ I(M; Y | M^0) - I(M; Z | M^0) \right] + 3\gamma
\]
where (a) follows from Csiszár’s sum identity \([8]\). Also, from (4),

\[
R \leq I(M; Y^{T-1}, Z_{T+1}^n; T) + I(M; Z_T^{T-1}, Z_{T+1}^n, M^n) + 2\gamma.
\]

Hence, from (4),

\[
R \leq I(M^n, Y^{T-1}, Z_{T+1}^n, T; Y_T) + I(M; Y_T^{T-1}, Z_{T+1}^n, M^n, T) - I(M; Z_T^{T-1}, Z_{T+1}^n, M^n) + 3\gamma.
\]

Finally, let \(V = (M^n, Y^{T-1}, Z_{T+1}^n, T), U = (V, M), X = X_T, Y = Y_T \) and \(Z = Z_T\). Then, clearly, \(V - U - X - (Y, Z)\). Substituting above gives the claimed theorem.

C. Discussions

1) Plausible deniability vs Secrecy: In the following discussion, we compare the capacity region \(\mathcal{R}_m\) to rate regions for two standard information-theoretic secrecy problems – the Wire-Tap Channel \([2]\) and Broadcast Channel with Confidential messages \([3]\) (see Figure 4). To this end, we first adapt the following definitions from \([2]\), \([3]\).

Definition 1 (Rate-Equivocation Region). For a channel \(P_{YZX}\), the rate-equivocation region \(\mathcal{R}_{eqv}\) is the set of all non-negative \((R, R_e)\) pairs such for any \(\varepsilon > 0\) and large enough blocklength \(n\), there exists a code for the Wire-Tap Channel problem (Figure 4a) with \(H(M) \geq nR\), \(Q_{M, Y|X, M} < \varepsilon\) and \(H(M|Z) \geq nR_e\).

Definition 2 (Sum Capacity with Confidential and Public messages). For a channel \(P_{YZX}\), the sum capacity region with confidential and public messages \(\mathcal{R}_{sc}\) is the set of all non-negative \((R_1, R_e)\) pairs with \(R_1 \geq R_e\) for which, given any \(\varepsilon > 0\), for a large enough blocklength \(n\), there exists a code for the Broadcast Channel with Confidential Messages setup (Figure 4b) with \(H(M_0) \geq n(R - R_1)\), \(H(M_1) \geq nR_1\), \(Q_{M_0, M_1, X|Y,Z} < \varepsilon\) and \(H(M_1|Z) \geq nR_1 - \varepsilon\).

We note that in the Message deniability setting, a code that is \((\varepsilon, R)\)-reliable and is accompanied with a \((\delta, D)\)-plausibly deniable faking procedure also achieves the rate pair \((R, D - O(\sqrt{\delta}))\) in the Wire-Tap setup. The following proposition makes this property precise.

Proposition 1. For a channel \(P_{YZX}\), let \(\mathcal{G}\) be an \((\varepsilon, R)\)-reliable code with message \(M\) and let \(M^{\varepsilon}\) be \((\delta, D)\)-plausibly deniable for \(M\). Then, there exists \(\mu\) depending only on \(P_{YZX}\) such that

\[
H(M|Z) \geq nD - n\mu \sqrt{\delta} - 2\delta.
\]

Proof: The above proposition is a direct consequence of Lemma 1. Specifically, note that there exists \(\lambda = \lambda(P_{YZX})\) such that

\[
H(M|Z) \geq H(M|Z) + I(M; Z|M^{\varepsilon}) - \delta - n\lambda \sqrt{\delta}
\]

\[
= H(M|Z) + H(M|M^{\varepsilon}) - H(M|Z, M^{\varepsilon}) - \delta - n\lambda \sqrt{\delta}
\]

\[
\geq H(M|M^{\varepsilon}) - \delta - n\lambda \sqrt{\delta}
\]

\[
= H(M^{\varepsilon}|M) + H(M) - H(M^{\varepsilon}) - \delta - n\lambda \sqrt{\delta}
\]

\[
\geq nD - 2\delta - n\lambda \sqrt{\delta}.
\]

The above proposition leads to the following corollary.

Corollary 1. \(\mathcal{R}_{sc} \subseteq \mathcal{R}_m \subseteq \mathcal{R}_{eqv}\).

Proof: The first inclusion, \(\mathcal{R}_{sc} \subseteq \mathcal{R}_m\), follows directly by comparing our characterization of \(\mathcal{R}_m\) with the capacity expression from \([3]\). Note that in the setting of \([3]\), the public message of rate \(R - R_1\) is intended to be decoded by both the receivers, while in our achievability, we require that it be decoded only by Bob. This allows us to operate with public message rates as high as...
Fig. 4: In the Wire-Tap Channel problem (first introduced by [2] and explored further in [3]), the goal for Alice is to transmit a confidential message \( m \) to the legitimate receiver Bob while ensuring that the “leakage” to the eavesdropper Judy (measured through the rate of equivocation) is smaller than a threshold. The capacity region for this problem (see Definition [1]) exhibits a tradeoff between the message rate \( R \) and the equivocation rate \( R_e \). The Broadcast Channel with Confidential Messages setup (introduced by [3]) generalizes the Wire-Tap Channel model to include a “public” message \( m_0 \) that is meant to be decoded by both Bob and Judy. Similarly to the Wire-Tap Channel, this setup also includes a confidential message \( m_1 \) that is meant to be decoded by only Bob while ensuring that the leakage to Judy is smaller than a threshold. In general, the capacity region for this setup exhibits a tradeoff between three parameters – the rate of the public message \( R_0 \), the rate of the confidential message \( R_1 \), and the equivocation rate. In our discussion, we only consider a two-dimensional projection of this region (see Definition [2]) to the set of \( (R_0, R_1) \) pairs that ensure that the equivocation about the message \( m_1 \) is arbitrarily close to the entropy of \( m_1 \). The reader is referred to [9] for an excellent introduction to these and other information-theoretic security problems.

The following example illustrates that both inclusions in the above corollary may be strict.

**Example 1 (Binary Erasure Eavesdropper).** Consider the example of Figure 1. Let \( X = Y = \{0, 1\} \), \( Z = \{0, E, 1\} \), and

\[
P_{YZ|X}(yz|x) = \begin{cases} 
1 - p & \text{if } (y, z) = (x, x), \\
p & \text{if } (y, z) = (x, E), \\
0 & \text{otherwise.}
\end{cases}
\]
For the message deniability problem, the capacity region $R_m$ is given by

\[ 0 \leq R \leq 1 \\
0 \leq D \leq \min \left\{ \frac{p(1-R)}{1-p}, R \right\}. \]

It is instructive to compare this region with the regions $R_{bc}$ and $R_{equiv}$. For this example, the Rate-Equivocation region consists of all $(R, R_e)$ pairs satisfying

\[ 0 \leq R \leq 1 \\
0 \leq R_e \leq \min\{p, R\}. \]

Next, the region $R_{bc}$ consists of all $(R, R_1)$ pairs satisfying

\[ 0 \leq R \leq 1 \\
0 \leq R_1 \leq \min \left\{ \frac{p(1-p-R)}{1-2p}, R \right\}. \]

Comparing the above regions, it is evident that the inclusion relation in Corollary 1 may be strict. The plot shown in Figure 5 compares these regions.

Fig. 5: Comparision of $R_m$ with $R_{bc}$ and $R_{equiv}$ in Example 1

2) Rate of deniability as the Equivocation rate: Even though we define the rate of deniability as an operational property of the faking procedure, surprisingly, it also has a rough interpretation as the rate of equivocation given the eavesdropper channel output as well as the fake message. This is especially interesting in light of Example 1 that shows that the rate of deniability may be strictly smaller than the equivocation rate at the eavesdropper in the Wire-Tap Channel setting. The following proposition states this property formally.

**Proposition 2.** Let $M^\omega$ be $(\delta, D)$-plausibly deniable for $M$ given $Z$. Then, there exists $\mu \geq 0$ depending only on $P_{YZIX}$ such that

\[ nD - \delta - n\mu \sqrt{\delta} \leq H(M|M^\omega, Z) \leq nD + \delta + n\mu \sqrt{\delta}. \]

**Proof:** Note that

\[ H(M|M^\omega, Z) = H(M|M^\omega) - I(M; Z|M^\omega) \]
\[ = H(M^\omega|M) - H(M^\omega) + H(M) - I(M; Z|M^\omega) \]
\[ = nD - H(M^\omega) + H(M) - I(M; Z|M^\omega). \]

Applying Lemma 1 and the non-negativity of mutual information to the terms on the left hand side above gives the claimed result.
V. Transmitter and Receiver Deniability

A. Zero Information Variables

For a random variable $W \sim P_W$ and a channel $P_{ZW}$, we define the following relation: for $w_1, w_2 \in \mathcal{W}$, we say that $w_1 \sim w_2$ if $P_{ZW}(z|w_1) = P_{ZW}(z|w_2)$, for all $z \in \mathcal{Z}$. It is evident that this is an equivalence relation. Let $\mathcal{W}_0$ represent the set of equivalence classes of this relation. We define the zero-information random variable $U_0$ of $W$ w.r.t. $P_{ZW}$ as a random variable taking values in $\mathcal{W}_0$ and is such that $W \in U_0$. For each $w \in \mathcal{W}$, we will call the corresponding $u_0$ its zero-information symbol. Clearly, $U_0$ is a function of $W$. Note that $U_0 - W - Z$ (since $U_0$ is a function of $W$), $W - U_0 - Z$ (by definition), and $P_{ZW}(z|w) = P_{ZWU_0}(z|w, u_0) = P_{ZU_0}(z|u_0)$ if $u_0$ is the zero-information symbol of $w$. Figure 6 shows an example of a zero-information variable.

![Diagram showing zero-information variables](image_url)

The above lemma shows that for achieving the rates claimed in Theorems 2 and 3, it suffices to restrict our auxiliary random variables to the corresponding zero information variables. This is critical to our proofs.
B. Transmitter Deniability

We begin our proof for Theorem 2 by stating two lemmas that lead to our converse arguments.

**Lemma 3.** Let \( X^o \) be \((\delta, D)\)-plausibly deniable for \( X \) given \( Z \) and \( X^o - X - Z \). Then, there exists a constant \( \kappa \) depending only on \( P_{Z|X} \) such that

\[
\begin{align*}
I(X; Z|X^o) &\leq nk \sqrt{\delta}, \\
|\mathbb{H}(X|X^o) - \mathbb{H}(X^o|X)| &\leq nk \sqrt{\delta}, \\
|\mathbb{H}(X) - \mathbb{H}(X^o)| &\leq nk \sqrt{\delta}, \text{ and} \\
|\mathbb{H}(X|X^o, M) - \mathbb{H}(X^o|X, \text{Msg}(X^o))| &\leq nk \sqrt{\delta}.
\end{align*}
\]

**Proof:** We explicitly only prove the first inequality. The other inequalities follow from a similar reasoning.

\[
I(X; Z|X^o) = \mathbb{H}(Z|X^o) - \mathbb{H}(Z|X) = \sum_{z \in Z, x \in X} Q_{Z,X^o}(z, x) \log \frac{Q_{X^o}(x)}{Q_{Z,X^o}(z, x)} - \log \frac{Q_X(x)}{Q_{Z,X}(z, x)}
\]

\[
\leq D(Q_{X^o}||Q_X) - D(Q_{Z,X^o}||Q_{Z,X}) + \sum_{z \in Z, x \in X} \left[ Q_{Z,X^o}(z, x) - Q_{Z,X}(z, x) \right] \log \frac{Q_X(x)}{Q_{Z,X}(z, x)}
\]

\[
\leq \delta + n \sqrt{2\delta} \max_{(z, x) : P_{Z|X}(z|x) > 0} \log \frac{1}{P_{Z|X}(z|x)}. \]

In the last step above, we use the fact that \( X^o \) is \((\delta, D)\)-plausibly deniable \( X \) for \( Z \). The bound on the first term follow from definition, the second from non-negativity of K-L divergence, while the last term is bounded by applying Pinsker’s inequality. ■

The following lemma follows from a standard chain of information inequalities with Lemma 3 as a starting point and single-letterizing the resulting expressions.

**Lemma 4.** Consider a broadcast channel \( P_{YZ|X} \) from Alice to Bob and Judy. Suppose that there exists an \((\epsilon, R)\)-reliable code \( C \) of blocklength \( n \), such that \( X^o \) is \((\delta, D)\)-plausibly deniable for \( X \) given \( Z \). Then, there exists \( \gamma = \gamma(\epsilon, \delta) > 0 \) satisfying \( \lim_{\epsilon, \delta \to (0, 0)} \gamma = 0 \) such that \( R \leq I(X; Y) + R \), and \( D \leq I(X; Y|U) + \gamma \), where the random variables \( U, X, Y, \) and \( Z \) satisfy \( U - X - (Y, Z) \) and \( I(X; Z|U) \) \( \leq \gamma \).

**Proof:** Note that \( Y - X - X^o \). We use Lemma 3 below.

\[
nD = \mathbb{H}(\text{Msg}(X^o)|X) = \mathbb{H}(X^o|X) - \mathbb{H}(X^o|X, \text{Msg}(X^o))
\]

\[
\leq \mathbb{H}(X|X^o) - \mathbb{H}(X|X, \text{Msg}(X)) + 2nk \sqrt{\delta}
\]

\[
\leq \mathbb{H}(X|X^o) - \mathbb{H}(X|Y, X^o, \text{Msg}(X)) + 2nk \sqrt{\delta}
\]

\[
= \mathbb{H}(X|X^o) - \mathbb{H}(X|Y, X^o) + I(X; \text{Msg}(X)|Y, X^o) + 2nk \sqrt{\delta}
\]

\[
\leq I(X; Y|X^o) + n\epsilon + 2nk \sqrt{\delta}
\]

\[
\leq I(X; Y|X^o) - I(X; Z|X^o) + n\epsilon + 3nk \sqrt{\delta}
\]

\[
= \sum_{i=1}^{n} \left[ I(X; Y|X^o, Y^{i-1}) - I(X; Z|X^o, Z^n_{i+1}) \right] + n\epsilon + 3nk \sqrt{\delta}
\]

\[
\leq \sum_{i=1}^{n} \left[ I(X; Y|X^o, Y^{i-1}, Z^n_{i+1}) - I(X; Z|X^o, Y^{i-1}, Z^n_{i+1}) \right] + n\epsilon + 3nk \sqrt{\delta}
\]

\[
= \sum_{i=1}^{n} \left[ H(Y|X^o, Y^{i-1}, Z^n_{i+1}) - H(Y|X^o, Y^{i-1}, Z^n_{i+1}) - H(Y|X^o, Y^{i-1}, Z^n_{i+1}) + H(Z|X, X^o, Y^{i-1}, Z^n_{i+1}) + H(Z|X, X^o, Y^{i-1}, Z^n_{i+1}) \right] + n\epsilon + 3nk \sqrt{\delta}
\]

\[
\leq \sum_{i=1}^{n} \left[ H(Y|X^o, Y^{i-1}, Z^n_{i+1}) - H(Y|X^o, Y^{i-1}, Z^n_{i+1}) - H(Y|X^o, Y^{i-1}, Z^n_{i+1}) + H(Z|X, X^o, Y^{i-1}, Z^n_{i+1}) + H(Z|X, X^o, Y^{i-1}, Z^n_{i+1}) \right] + n\epsilon + 3nk \sqrt{\delta}.
\]

In the above, (a) and (c) follow from Lemma 3, (b) is a consequence of Fano’s inequality, (d) is an application of Csiszár’s sum identity, and (e) relies on the memoryless nature of the channel to argue that \( (Y_i, Z_i) - X_i - (X^o, Y^{i-1}, Z^n_{i+1}) \) is a Markov chain.
Next, we let \( U_i \equiv (X^o, Y^{i-1}, Z_{t+1}^n) \), and let \( T \) be a random variable independent of \((M, X^o, X, Y, Z)\) that is uniformly distributed over \([1 : n]\). Note that \( U_i - X_i - (Y, Z_i) \) is a Markov chain. The above inequalities are continued further as

\[
\begin{align*}
nD & \leq \sum_{i=1}^{n} \left[ \mathcal{H}(Y | U_i) - \mathcal{H}(Y | X_i, U_i) - \mathcal{H}(Z | U_i) + \mathcal{H}(Z | X_i, U_i) \right] + n\epsilon + 3n\kappa \sqrt{\delta} \\
& \leq nI(X_T; Y_T | U_T, T) - nI(X_T; Z_T | U_T, T) + n\epsilon + 3n\kappa \sqrt{\delta}.
\end{align*}
\]

(5)

Next, note that

\[
I(X_T; Z_T | U_T, T) = \frac{1}{n} \sum_{i=1}^{n} I(X_i; Z_i | X^o, Y^{i-1}, Z_{t+1}^n)
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \left[ \mathcal{H}(Z_i | X^o, Y^{i-1}, Z_{t+1}^n) - \mathcal{H}(Z_i | X_i, X^o, Y^{i-1}, Z_{t+1}^n) \right]
\]

\[
\leq \frac{1}{n} \sum_{i=1}^{n} I(X_i; Z_i | X^o, Z_{t+1}^n)
\]

\[< \kappa \sqrt{\delta}.\]

In the above, (a) follows by noting that for each \( i, Z_i - X_i - (X^{i-1}, Y^n_i, Z_{t+1}^n, Y^{i-1}, Z_{t+1}^n) \) is a Markov chain due to the memoryless nature of the channel \( P_{Z|X} \) and (b) follows from Lemma 3. Defining random variables \((U, X, Y)\) with \( Q_{U|X,Y}(u,x,y) = Q_{(U,T),X,Y}(u,x,y) \), we obtain

\[
D \leq I(X; Y | U) + \epsilon + 3\kappa \sqrt{\delta}.
\]

Notice that \( Q_{Z|X} \) is the same as the channel transition probability \( P_{Z|X} \). Further, \( U - X - (Y, Z) \) is a Markov chain and \( I(X : Z | U) < \kappa \sqrt{\delta} \). Thus, \( U \) satisfies the constraints from the lemma statement. Finally, we bound the rate as follows.

\[
nR = \mathcal{H}(M) \leq I(X; Y) + n\epsilon
\]

\[
\leq \sum_{i=1}^{n} I(X_i; Y_i) + n\epsilon
\]

\[
= nI(X_T; Y_T | T) + n\epsilon
\]

\[
\leq nI(T, X_T; Y_T) + n\epsilon
\]

\[
= nI(X_T; Y_T) + nI(T; Y_T | X_T) + n\epsilon
\]

\[< nI(X_T; Y_T) + n\epsilon.
\]

In the above, (a) follows from Fano’s inequality and (b) from the fact that \( Q_{Y_T|X_T,Y}(y|x,t) = P_{Y|X}(y|x) \). Setting \( \gamma = \epsilon + 3\kappa \sqrt{\delta} \) proves the lemma.

We are now ready to formally prove Theorem 2.

**Proof of Theorem 2.** The converse for Theorem 2 follows from Lemma 4 by applying standard continuity arguments from Lemma 3 to show that the bounds on \( R \) and \( D \) satisfy the theorem statement as \( \delta \) vanishes.

We now give a proof sketch for the achievability of claimed rate region. Our achievability uses a superposition code for the broadcast channel \( P_{X|Z} \). To this end, choose random variables \((X, U)\) satisfying the conditions in the theorem with \( U \) as the zero information variable of \( X \) w.r.t. \( P_{ZX} \). Recall that Lemma 2 guarantees that there is no loss of optimality in choosing \( U \) as the zero information variable of \( X \) w.r.t. \( P_{ZX} \).

We consider a two layer random code. For any \( \epsilon > 0 \), first, \( \gamma = \{u(1), \ldots, u(2^n R(X,U)-\epsilon)\} \) is generated by drawing each \( u_t(j) \) independently from the distribution \( P_U \). Next, for each \( u \in \gamma \), a sub-code \( \gamma_u = \{x(u,1), \ldots, x(u,2^n R(X,Y,U)-\epsilon)\} \) is generated by drawing each \( x_l(u, j) \) independently from the distribution \( P_{X|U}(\cdot | u) \). The codebook \( \gamma = \{x^{\gamma}(m) : m \in \mathcal{M}\} \) is formed by taking...
the union $\cup_{a \in \mathcal{C}} \mathcal{G}_u$. Finally, the faking procedure simply accepts the transmitted codeword (say, $x$) and outputs a uniformly drawn codeword from the sub-code that contains $x$ (say, $\mathcal{C}_u$).

Since the reliability of the above code follows from standard arguments for superposition coding (see [8] for example), we skip the detailed analysis here. The plausible deniability for the code follows directly from the construction by noting that for every $x \in \mathcal{C}_u$, $u$ is precisely the sequence of the zero information symbols of $x$ w.r.t. $P_{Z|X}$. Thus, for any $x, x' \in \mathcal{C}_u$ and $z \in \mathcal{Z}^n$, $Q_{XZ}(x|z) = Q_{XZ}(x'|z)$.

### C. Receiver Deniability

In this section, give the proof of our achievability for the Receiver Deniability setting and comment on the tightness of this region in subsequent remarks.

**Proof of Theorem 3** Let $\epsilon > 0$ and set

$$R = I(X; Y) - \rho$$

for some $\rho > \epsilon$. Consider the following codebook and faking procedure.

**a) Codebook generation:** The codebook $\mathcal{C} = \{x^{(\epsilon)}(m) : m \in \mathcal{M}\}$ is generated by drawing each $x^{(\epsilon)}(m)$ independently from the distribution $P_X$. Let $Pr$ be the probability distribution over the random generation of the codebook.

**b) Decoding:** Upon receiving $y$, the decoder looks for $m \in \mathcal{M}$ such that $(x^{(\epsilon)}(m), y) \in A_1(\alpha)(X, Y)$.

**c) Faking procedure:** Given $y$, the faking procedure first generates the unique $v$ where, for each $i$, $v_i$ represents the zero information symbol of $y_i$ w.r.t. $P_{Z|Y}$. Next, $Y^m$ is drawn from $\mathcal{Y}^m$ according to the conditional distribution $Q_{Y|Y^m} = Q_{Y|V}$.

**d) Analysis:** Note that $Y^m - V = (Y, X, Z)$ is a Markov chain. For similar reasons as in transmitter deniability, these ensure that the parameter $\delta$ is zero. To this end, we first observe that for any $(y, v, z) \in \mathcal{Y}^m \times \mathcal{Y}^m \times \mathcal{Z}^n$,

$$Q_{YVZ}(y, v, z) = Q_{YV}(y, v)Q_{Z|Y}(z|v)$$

and

$$Q_{Y^mVZ}(y, v, z) = \sum_{y' \in \mathcal{Y}^m} Q_{Y^{m_0}VZ}(y', y, v, z)$$

$$= \sum_{y' \in \mathcal{Y}^m} Q_{YV}(y'|v)Q_{Y^m|V}(y|v)Q_{V|Y}(v|z|v')$$

$$= Q_{Y^mV}(y'|v)Q_{V|Y}(v|z|v')P_{Z|Y}(z|v)$$

$$= Q_{Y^mV}(y|v)Q_{V|Y}(v|z|v)$$

$$= Q_{Y^mVZ}(y, v, z).$$

In the above, $(a)$ and $(d)$ follow from the dependence structure of the random variables $Y, Y^m, V$, and $Z$, $(b)$ is a consequence of the channel being stochastically degraded, $(c)$ and $(e)$ are true since $V$ is the zero information variable of $Y$ w.r.t. $P_{Y|Z}$, and $(f)$ is implied by the faking procedure used to generate $Y^m$. Thus,

$$\delta = D(Q_{Y^m|Z}||Q_{YZ})$$

$$\leq D(Q_{Y^mVZ}||Q_{YVZ})$$

$$= 0.$$

Next, we analyze the rates $(R, D)$ that our code and faking procedure can achieve. Let $\alpha \in (0, 1)$. The reliability analysis is similar to Shannon’s channel coding theorem. Let $\mathcal{G}_1 = \{C : C_{MXY}(M \neq \hat{M}) < \epsilon\}$ denote the class of codebooks that have an average error probability smaller than $\epsilon$. Following the standard proof of reliability of random codes, there exists $n_1 = n_1(\alpha)$ such that as long as $R < I(X; Y)$ and $n > n_1$,

$$Pr(\mathcal{G}_1) \geq 1 - \alpha/4.$$(7)
In the following we assume that $C \in G_1$ and prove that, with a high probability over the codebook generation, the random code also admits a faking procedure that satisfies the plausible deniability requirement. To this end, the following chain of inequalities give a lower bound on $D$ for the code $C$.

$$
D = H(D_{\text{Dec}}(Y^n)|Y) \\
= H(D_{\text{Dec}}(Y)|V) \\
\geq I(D_{\text{Dec}}(Y); X|V) \\
= I(D_{\text{Dec}}(Y), Y; X|V) - I(X; Y|V, D_{\text{Dec}}(Y)) \\
= I(X; Y|V) - I(X; Y|V, D_{\text{Dec}}(Y)) \\
\geq I(X; Y|V) - ne.
$$

(8)

In the above, (8) follows from the faking procedure inducing $Q_v^n|Y = Q_v^n|Y$, the fact that $V$ is a function of $Y$, and the Markov chain $Y^n - V - Y$. Fano’s inequality implies (9) (assuming that $C \in G_1$). Note that the above bound is a multi-letter bound that depends on the specific codebook $C$. A single letter bound depending only on the probability distribution of the single letter random variables follows from concentration arguments over the codebook generation process. In the following, we argue that, with high probability over the generation of $C$, $I(X; Y|V) \geq nH(X; Y|V) - ne$ for a large enough $n$. For every $v \in \mathcal{V}$, let the multi-set $C_v \triangleq \{x \in C : (x, v) \in A_e^n(X, Y)\}$. Further, for every $x \in \mathcal{X}^n$, let $\mathcal{M}_x \triangleq \{m \in \mathcal{M} : x^{(e)}(m) = x\}$. First, note that

$$
I(X; Y|V) \geq H(X|V) - ne.
$$

(10)

by Fano’s inequality (assuming that $C \in G_1$). Then, given a code $C$, there exists $e' = e'(e)$ satisfying $\lim_{e \to 0} e' = 0$ and

$$
H(X|V) \geq \sum_{(x,v) \in A_e^n(X,V)} Q_{X,V}(x,v) \log_2 \frac{Q_v(v)}{Q_{X,V}(x,v)} \\
= \sum_{(x,v) \in A_e^n(X,V)} Q_X(x)P_{V|X}(v|x) \log_2 \frac{\sum_{x' \in C} 2^{-nR} P_{V|X}(v|x')} {Q_X(x)P_{V|X}(v|x)} \\
\geq \sum_{(x,v) \in A_e^n(X,V)} Q_X(x)P_{V|X}(v|x) \log_2 \frac{\sum_{x' \in C_v} P_{V|X}(v|x')} {\mathcal{M}_x} P_{V|X}(v|x) \\
\geq \sum_{(x,v) \in A_e^n(X,V)} Q_X(x)P_{V|X}(v|x) \log_2 \frac{\mathcal{M}_x}{\mathcal{M}_x} - ne'.
$$

(11)

In the above, (a) is obtained by expressing $Q_X(x)$ as $2^{-nR} \mathcal{M}_x$. (b) follows by noting that for every $(x, v)$ belonging to $A_e^n(X, V)$, $\log_2 \frac{\mathcal{M}_x}{\mathcal{M}_x} - nH(V|X) < ne'$ for some $e' > 0$ that can be made arbitrarily close to 0 as $e$ approaches 0. We now show that, with high probability over the random generation of $C$, the expression in (11) is lower bounded in the desired manner. To this end, define the following three desirable events over the codebook generation process.

$$
G_2 \triangleq \left\{ C : \sum_{(x,v) \in A_e^n(X,V)} Q_X(x)P_{V|X}(v|x) > (1 - \epsilon) \right\} \\
G_3 \triangleq \left\{ C : |\mathcal{M}_x| \geq 2^{nR - I(X; V) - e'} \forall v \in A_e^n(V) \right\} \\
G_4 \triangleq \left\{ C : |\mathcal{M}_x| < 2^ne \forall x \in A_e^n(X) \right\}.
$$

In the above, $\epsilon''$ is a constant that is specified later. Note that if $C \in \cap_{i=1}^4 G_i$, then Eq. (9)-(11) imply that

$$
D \geq \frac{1}{n} \left( (1 - \epsilon) \log_2 \frac{2^{nR - I(X; V) - e''}} {2^ne} - 2(\epsilon + e') \right) \\
= \left( (1 - \epsilon)(I(X; Y|V) - \rho - \epsilon - \epsilon'' - 2(\epsilon' + \epsilon')) \right) \\
\geq I(X; Y|V) - (\rho + 3\epsilon + \epsilon log_2 |\mathcal{X}| + 2\epsilon' + \epsilon'').
$$

(12)

Recall that $C$ is a multi-set with possibly repeated elements. As a result, $C_v$ may also contain codewords that have multiplicity greater than 1.
We next lower bound the probabilities of each of the above events.

**i) Event $G_2$:** First observe that

$$
\sum_{(x,v) \in A^n_\epsilon(X,V)} Q_X(x)P_{V|X}(v|x) \geq \frac{|\mathcal{E}\cap A^n_\epsilon(X)|}{|\mathcal{E}|} \min_{x \in A^n_\epsilon(X)} P_{V|X}(v \in A^n_\epsilon(V|x|x)).
$$

To bound the right hand side above, we first note that using the additive form of the Chernoff bound and the definition of strong typicality,

$$
E_{\mathcal{E}} \left[ \frac{|\mathcal{E}\cap A^n_\epsilon(X)|}{|\mathcal{E}|} \right] = \Pr(x \in A^n_\epsilon(X)) \geq 1 - \frac{1}{|\mathcal{X}|} \max_{x \in \mathcal{X}} \Pr_X \left( \left| \frac{|\{i : x_i = x\}|}{n} - \Pr_X(x) \right| > \frac{\epsilon}{|\mathcal{X}|} \right) \\
\geq 1 - |\mathcal{X}| \exp \left( - \frac{n \epsilon^2 \min_{x \in \mathcal{X}} P_X(x)}{4|\mathcal{X}|^2} \right).
$$

In particular, for a large enough $n$, we have

$$
E_{\mathcal{E}} \left[ \frac{|\mathcal{E}\cap A^n_\epsilon(X)|}{|\mathcal{E}|} \right] \geq 1 - \epsilon/4.
$$

Next, by standard properties of the conditionally typical set, we have, for large enough $n$,

$$
P_{V|X} \left( v \in A^n_\epsilon(V|x) \right) \geq 1 - \epsilon/4.
$$

Combining (14) and (15), we conclude that there exists $n^*$ such that for every $n > n^*$,

$$
E_{\mathcal{E}} \left[ \frac{|\mathcal{E}\cap A^n_\epsilon(X)|}{|\mathcal{E}|} \right] \geq \min_{x \in A^n_\epsilon(X)} P_{V|X}(v \in A^n_\epsilon(V|x|x)) > 1 - \frac{\epsilon}{2}.
$$

The above expression gives, in expectation, a lower bound on the left hand side of (13). A further concentration argument over the i.i.d. generation of the codebook shows the existence of $n_2 = n_2(\epsilon)$ such that whenever $n > n_2(\alpha)$,

$$
\Pr(G_2) \geq 1 - \alpha/4.
$$

**ii) Event $G_3$:** Next, note that for any $v \in A^n_\epsilon(V)$, there exists $n^\#$ and $\epsilon'' = \epsilon''(\epsilon)$ satisfying $\lim_{n \to 0} \epsilon'' = 0$ and

$$
E[|\mathcal{E}_v|] = 2^{n^\#} P_X((x, v) \in A^n_\epsilon(X, V)) \\
\geq 2^{(nR-H(X,V)-\epsilon''/2)}
$$

whenever $n > n^\#$. Now, since each codeword falls in $\mathcal{E}_v$ in an independent and identical manner over the codebook generation, the true value of $\mathcal{E}_v$ concentrates around its mean with a high probability. In particular, by applying Chernoff bound on $|\mathcal{E}_v|$, we obtain that there exists $n_3 = n_3(\epsilon)$ such that for every $n > n_3(\alpha)$,

$$
\Pr(G_3) \geq \Pr(|\mathcal{E}_v| \geq 2^{-\alpha^2/2} E[|\mathcal{E}_v|]) > 1 - \alpha/4.
$$

**iii) Event $G_4$:** Finally, let $\beta = 2^{n^\#}$, and observe that there exists $\epsilon''' = \epsilon'''(\epsilon)$ such that $\lim_{n \to 0} \epsilon''' = 0$ and

$$
\log_2 \left( \Pr(\mathcal{E} \notin G_4) \right) \leq \log_2 \sum_{\mathcal{E} \notin \mathcal{A}^n_\epsilon(X)} \sum_{x \in A^n_\epsilon(X)} \prod_{m \in \mathcal{M}} \Pr(x^{(m)}(1) = x) \\
\leq \log_2 \left( \frac{1}{\beta} \right) + \log_2 \sum_{x \in A^n_\epsilon(X)} \left( \Pr(\mathcal{E}^{(x)}(1) = x) \right)^\beta \\
\leq\beta \log_2 \left( \frac{1}{\beta} \right) + \log_2 |A^n_\epsilon(X)| + \beta \log_2 \max_{x \in A^n_\epsilon(X)} \Pr(x^{(x)}(1) = x) \\
\leq \beta \log_2 \left( \frac{1}{\beta} \right) + \log_2 |\mathcal{M}| - \log_2 |\mathcal{M}| \leq (\beta - 1)nH(X) + (\beta + 1)n\epsilon'''
$$
\[ R \leq H(1) \leq 2^n(R - \epsilon) + \log_2 e - (2^{n\epsilon} - 1)nH(X) + (2^{n\epsilon} + 1)n\epsilon'\prime'\prime \]
\[ = 2^n(R - H(X) - \epsilon + \epsilon'\prime'\prime) + \log_2 e + n(H(X) + \epsilon'\prime'\prime) \]
\[ \leq 2^n(R - H(X) - \epsilon + \epsilon'\prime'\prime) + \log_2 e + n(H(X) + \epsilon'\prime'\prime) \]
\[ = 2^n(n(-\rho - \epsilon + \epsilon'\prime'\prime) + \log_2 e + n(H(X) + \epsilon'\prime'\prime). \]  

(18)

In the above, (a) is a standard upper bound on \( \left\lceil \frac{b}{\rho} \right\rceil \) in terms of the binary entropy function \( H_b(\rho/|A|) \). (b) obtained by noting that there exists \( \epsilon'\prime'\prime \) such that for every \( \epsilon > 0 \), \( |A|^n(X) \leq 2^{nH_2(X) + \epsilon} \) and \( P_X(x) \leq 2^{-n(H(X) + \epsilon'\prime'\prime)} \) for each \( x \in A^n(X) \). Lastly, (c) is obtained by using the fact that for every \( a > 0 \), \( \log_2 a = \log_2 e \ln a \leq (a-1)\log_2 e \). Note that as long as \( \rho \) is strictly greater than \( \epsilon'\prime'\prime - \epsilon \), the right hand side of (18) diverges to \(-\infty\) as \( n \) increases without bound. In particular, this implies that there exists \( n_4 \) such that for every \( n > n_4(\alpha) \),

\[ \Pr(G_4) > 1 - \alpha/4. \]  

(19)

Finally, combining (7), (16), (17), and (19) we conclude that, whenever \( n > \max\{n_1, n_2, n_3, n_4\} \), with probability at least \( 1 - \alpha \), the randomly drawn code is simultaneously \((\epsilon, R)\)-reliable and \((0, D)\)-plausible deniable where \((R, D)\) satisfy the lower bounds in (6) and (12). Since \( \rho \) and \( \epsilon \) can be made arbitrarily close to zero, this shows the achievability of all rates in the interior of the claimed region.

Remark 1. Unlike the transmitter deniability setting, the rate region stated in Theorem 3 may not equal the capacity region. For instance, introducing an auxiliary variable \( U \) such that \( U - X \rightarrow (Y, Z) \) and performing superposition coding may increase the achievable rate region. This leads to the following proposition.

**Proposition 3.** Let \( P_{Z|Y}(z|y) > 0 \) for all \((z, y)\). Then, \( \mathscr{R}_y \) includes all \((R, D)\) pairs satisfying

\[ 0 \leq R \leq I(X; Y), \text{ and } 0 \leq D \leq I(X; Y|U, V) \]

for some \((U, V)\) s.t. \( U - X \rightarrow (Y, Z) \), \( Y - (U, V) \rightarrow Z \), and \( V - (U, Y) \rightarrow (X, Z) \).

The above proposition can be proved using superposition coding (see [3], for example). Alice first generates a “cloud-centre” codeword \( U \) drawn from the distribution \( P_U \) and subsequently generates a “satellite” code \( X \) for each \( U \) following the distribution \( P_{X|U} \). Bob first decodes \( U \), computes the zero information sequence \( V \) w.r.t. the conditional distribution \( P_{Z|X,U} \), and outputs \( Y' \) using the conditional distribution \( Q_{Y'|U,V} = Q_{Y|U,V} \). Note that choosing \( U \) to be a constant gives the rate region of Theorem 3.

Remark 2. The requirement \( P_{Z|Y}(z|y) > 0 \) in Proposition 3 appears to be a technical artefact of our plausibility metric being defined in terms of the Kullback-Leibler divergence. Specifically, for the scheme outlined in Proposition 3, this condition ensures that \( Q_{Z,Y} \) is absolutely continuous w.r.t. \( Q_{Y|V} \), and hence, the divergence can be bounded even when there is non-zero probability of Bob incorrectly decoding \( U \). It is conceivable that using a different plausibility metric such as the variational distance \( \frac{1}{2}\|Q_{Y,Z} - Q_{Y|V}Z\|_1 \) eliminates this requirement.

**D. Discussions**

1) An example:

**Example 2.** Consider the channel with \( \mathcal{X} = \mathcal{Y} = \mathcal{Z}, \ Y = X \) and \( P_{Z|X} \) as in Figure 6 i.e.,

\[ P_{Z|X}(y, z|x) = \begin{cases} 0.3 & (x, y, z) \in \{(1, 1, 1), (2, 2, 1)\} \\ 0.7 & (x, y, z) \in \{(1, 1, 2), (2, 2, 2)\} \\ 0.4 & (x, y, z) = (3, 3, 2) \\ 0.6 & (x, y, z) = (3, 3, 3) \\ 0 & \text{otherwise.} \end{cases} \]

We characterize the capacity region \( \mathcal{R}_y \) by restricting our choice of the auxiliary random variable \( U \) to the zero-information random variable. For the above conditional distribution, the zero-information random variable of \( X \) w.r.t. \( P_{Z|X} \) takes two values \(-u_1 = (1, 2)\) and \(-u_2 = (3)\). Since \( X = Y, I(X; Y) = H(X) \) and \( I(X; Y|U) = P_X(1) \log_2 \frac{P_{X|U}(1)}{P_{X|U}(2)} - P_X(2) \log_2 \frac{P_{X|U}(2)}{P_{X|U}(1)} \). The capacity region \( \mathcal{R}_y \) (Figure 7) consists of all \((R, D)\) pairs satisfying the following

\[ D \leq R \leq H_f(\frac{D}{2}, \frac{D}{2}, 1 - D) \]
\[ 0 \leq D \leq 1, \]

where \( H_f(\cdot, \cdot, \cdot) \) represents the ternary entropy function. Interestingly, the capacity region depends on the conditional distribution \( P_{Z|X} \), only through the zero-information variable induced by it – all conditional distributions \( P_{Z|X} \) that induce the same zero-information variable have the same capacity region (assuming \( P_{Y|X} \) is unchanged). This is a general feature of capacity regions for the transmitter deniability problem.
2) Rate of deniability as the Equivocation rate: Similar to the Message Deniability setting, we can attach a secrecy interpretation to the rate of deniability for faking procedures that are plausibly deniable. The following proposition mirrors Proposition 2.

**Proposition 4.** Let $X^v$ be $(\delta, D)$-plausibly deniable for $X$ given $Z$ respectively. Then, there exists $\mu \geq 0$ depending only on $P_{Z|X}$ such that

$$nD - n\mu \sqrt{\delta} \leq H(M|Z, X^v) \leq nD + n\mu \sqrt{\delta}.$$  

**Proof:** The proof relies on the Lemma 3 and proceeds in similar spirit as Proposition 2. To this end, let $\kappa$ be the constant defined in Lemma 3. Note that

$$H(M|Z, X^v) = H(M|X^v) - I(M; Z|X^v)$$

$$= nD + H(M|X^v) - H(\text{msg}(X^v)|X) - I(M; Z|X^v).$$

Applying the non-negativity of mutual information and Lemma 3 we obtain

$$H(M|Z, X^v) \leq nD + H(M|X^v) - H(\text{msg}(X^v)|X) \leq nD + n\kappa \sqrt{\delta},$$

and

$$H(M|Z, X^v) \geq nD - 2n\kappa \sqrt{\delta}.$$ 

Choosing $\mu = 2\kappa$ completes the proof. 

---

VI. **Concluding remarks**

In this paper, we have considered three different models of Plausible Deniability and give achievable rates for each model while also giving tight converses for the message deniability and transmitter deniability settings. It is evident that, at the very least, each capacity region is a subset of the Rate-Equivocation region. Intuitively, this may be interpreted as follows – any code that has a rate of deniability $D$ has the property that the equivocation at the eavesdropper is at least $D$ (otherwise, with high probability, the eavesdropper can detect a fake response). On the other hand, it is not a priori clear whether the achievable rates for any one model considered in this paper is a subset of another – part of the difficulty in comparing the different settings arises from the fact that in each setting, the faking procedure accepts different inputs to generate the fake output.

A potential drawback of our Transmitter and Receiver Deniability results is that non-zero rates are possible only when non-trivial zero information variables exists. We note that the existence of such variables is guaranteed only for fairly special classes of channels – even for channels such as Binary Symmetric Channels, the only zero information variables are the channel inputs themselves. Further, the existence of non-trivial zero information variables may be rather fragile with respect to even slight perturbations in the channel conditional probability. However, even where non-zero rates are not possible, an asymptotically vanishing rate of communication may still possible over some channels, perhaps similarly to the “square-root law” observed in covert communications [10]–[13].

---

**References**

[1] J. Katz and Y. Lindell, *Introduction to Modern Cryptography.* Chapman & Hall/CRC, 2007.

[2] A. Wyner, “The Wire-tap Channel,” *Bell System Technical Journal, The,* vol. 54, no. 8, pp. 1355–1387, Oct 1975.

[3] I. Csiszar and J. Korner, “Broadcast Channels with Confidential Messages,” *IEEE Transactions on Information Theory,* vol. 24, no. 3, pp. 339–348, May 1978.
Appendix

Lemma 5. Let $\mathcal{P}$ be a compact subset of the set of probability measures over a finite set $\mathcal{B}$. Let $L: \mathcal{P} \rightarrow \mathbb{R}^+$ and $M: \mathcal{P} \rightarrow \mathbb{R}$ be functionals that are continuous with respect to the variational distance such that $L^{-1}(\{0\}) \neq \emptyset$. Then,

$$\lim_{\delta \to 0} \max_{P_B \in \mathcal{P} : L(P_B) < \delta} M(P_B) = \max_{P_B \in \mathcal{P} : L(P_B) = 0} M(P_B).$$

Proof: Recognizing that the left hand side is at least as large as the right hand side, we only show that the limit point of the right hand side cannot be larger than the right hand side.

To this end, let $M' = \lim_{\delta \to 0} \max_{P_B \in \mathcal{P} : L(P_B) < \delta} M(P_B)$. Thus, there exists a sequence $\{P_B^{(i)}\}_{i \in \mathbb{N}}$ in $\mathcal{P}$ such that $L(P_B^{(i)}) < 1/i$ and $\lim_{i \to \infty} M(P_B^{(i)}) = M'$. As $\mathcal{P}$ is a compact set under the variational distance, $\{P_B^{(i)}\}_{i \in \mathbb{N}}$ contains a subsequence $\{P_B^{(i)}\}_{i \in \mathbb{N}}$ that converges (in variational distance) to a limiting distribution $P_B'$. By continuity of $L$, we have

$$0 \leq L(P_B') = \lim_{i \to \infty} L(P_B^{(i)}) \leq \lim_{i \to \infty} 1/i = 0.$$

Thus, $M' = M(P_B') \leq \max_{P_B \in \mathcal{P} : L(P_B) = 0} M(P_B)$. 

\[\square\]