

Reliable, Deniable, and Hidable Communication over Parallel Link Networks

Swanand Kadhe, Sidharth Jaggi, Mayank Bakshi, and Alex Sprintson

Abstract

We consider the scenario wherein Alice wants to (potentially) communicate to the intended receiver Bob over a network consisting of multiple parallel links in the presence of a passive eavesdropper Willie, who observes an unknown subset of links. A primary goal of our communication protocol is to make the communication “deniable”, *i.e.*, Willie should not be able to *reliably* estimate whether or not Alice is transmitting any *covert* information to Bob. Moreover, if Alice is indeed actively communicating, her covert messages should be information-theoretically “hidable” in the sense that Willie’s observations should not *leak any information* about Alice’s (potential) message to him – our notion of hidability is slightly stronger than the notion of information-theoretic strong secrecy well-studied in the literature, and may be of independent interest. It can be shown that deniability does not imply either hidability or (weak or strong) information-theoretic secrecy; nor does any form of information-theoretic secrecy imply deniability. We present matching inner and outer bounds on the capacity for deniable and hidable communication over *parallel-link networks*.

I. INTRODUCTION

THE urge to communicate, to speak and be heard, is a fundamental human need. However, embedded within our increasingly sophisticated communication networks, Big Brother is often watching. There are situations where even the fact that communication is happening (not just the content of that communication), can have real-world consequences. For instance, if you are a politically active citizen in an authoritarian society with broad censorship powers, the mere fact that you are communicating with the outside world can be construed by those authorities as sufficient justification for reprisals.

The goal of this paper is to investigate a class of communication models with a threefold objective. Firstly, all communication from the source (Alice) to the destination (Bob) should be reliable, *i.e.*, Bob should be able to identify the messages intended for him and decode them with high accuracy. Secondly, if the communication is overheard by a third party (Willie), it should be deniable from Willie. That is, Willie should not even be able to reliably decide whether or not Alice is indeed communicating with Bob. Finally, the communication should be hidable from Willie, *i.e.*, the eavesdropper Willie should not be able to learn anything about Alice’s messages to Bob. Throughout the paper, we assume that Alice and Bob do *not* share any secret information that is not known to Willie.

Specifically, we consider the model wherein Willie is aware of Alice’s “innocent” communication patterns (when she is not communicating covertly with Bob). However, due to resource limitations, Willie cannot wiretap on all of Alice’s communication links, but only some of them. Willie, then, wishes to estimate whether Alice’s communication is routine, or malevolent. Furthermore, Willie is also interested in inferring some information about Alice’s (potential) covert communication with Bob.

For example, perhaps during the course of a workday, highly placed government official Alice sends text messages, makes phone calls, writes letters, and posts on various websites. Intelligence analyst Willie, who is suspicious that Alice is perhaps a foreign spy, is keeping tabs on some of this activity, but, critically, is unable to see all of it. Willie’s first goal is to differentiate between an “innocent” Alice (who communicates, in a somewhat predictable manner, with the outside world as a part of her daily life)

Swanand Kadhe and Alex Sprintson are with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, 77840 USA; e-mails: {kswanand1, spalex}@tamu.edu.

Sidharth Jaggi and Mayank Bakshi are with the Department of Information Engineering, The Chinese University of Hong Kong, New Territories, HK; emails: jaggi@ie.cuhk.edu.hk, mayank@inc.cuhk.edu.hk

and an “active” Alice (who is deliberately leaking secrets to spymaster Bob who gets to observe Alice’s transmissions on all channels). Willie’s second goal is to infer some information about the secrets that Alice might be leaking to Bob.

It is important to note that the conventional means of achieving secrecy, like cryptographic security, are not sufficient to achieve the goal of deniability. On the contrary, if Willie finds that the data being communicated is encrypted, it can arouse his suspicion that the covert communication is under progress.

Essentially, to be deniable, an active Alice has to look innocent in any subset of her communication links (since she does not know which of her communications might be wiretapped by suspicious authorities), *i.e.*, her communication on any subset of channels should be commensurate with what a normal person in Alice’s position would do. Her covert communication with Bob, then, must be a function of all these channels. The challenge for Alice is to be able to embed meaningful information to Bob on channels which individually look innocent (deniability), and in such a way that Willie can infer absolutely nothing about her covert communication with Bob (hidability).

We begin by formulating a mathematical model for this scenario, considering that Alice tries to communicate with Bob over a network with multiple parallel links, referred to as *parallel-link network*. Then, we formally define the conditions for deniability and hidability. Our condition of hidability is slightly stricter than that of strong information-theoretic secrecy.¹ We prove that the hidability always implies the strong information-theoretic secrecy, but the converse is not necessarily true. In particular, to show that the hidability condition is stricter, we give an example of a strongly information-theoretically secure scheme, which allows Willie to infer substantial information about an asymptotically large subset of covert messages with significantly high probability (even though the fraction of such feeble messages is asymptotically small).

At first glance, one might get tempted to argue that the deniability condition, which prevents Willie from estimating whether or not the cover communication is underway, will also restrict Willie from inferring any information about the (potential) covert communication. On the other hand, one can also think that a hidable scheme, which allows Willie to infer nothing about the covert communication, will inherently be deniable. However, we show, through examples, that a deniable scheme may not necessarily be hidable, and vice-versa.

Finally, we characterize the capacity for deniable and hidable communication over a parallel-link network.

Our main contributions can be summarized as follows. For a parallel-link network,

- We characterize the capacity for deniable communication, and present a stochastic encoding scheme which achieves the capacity.
- We characterize the capacity for deniable *and* hidable communication, and present a stochastic encoding based achievable scheme.

The information-theoretic techniques that we use enable us to attain separability between the deniable encoding and the hidable encoding. This essentially means that, by appropriately selecting one of the proposed schemes, the communication can be made either deniable (but not hidable) or hidable (but not deniable), or both deniable and hidable.

II. RELATED WORK

Several notions regarding the covert communication have been considered in the literature. Typically, in practical systems, *cryptographic techniques* are employed to achieve secrecy. However, even though cryptography allows communication at a rate higher than that possible by using other techniques, cryptographic techniques are not inherently deniable. Essentially, most cryptographic techniques make it computationally hard for an eavesdropper to distinguish the output (of the cryptographic system) from the output of a uniformly random noise sequence. Therefore, to achieve deniability, the outputs of most cryptographic schemes would still need to be “shaped” via techniques described in this paper.

¹A generalized form of the notion of hidability, referred to as *superstrong secrecy*, is proposed in [Chung-SSS].

The notion of *information-theoretic secrecy* has received a considerable amount of attention, see for example, [1], [2]. Ostensibly, the proposed notions of deniability and hidability seem to have significant overlap with the notions of information-theoretic secrecy. However, as we show in appendix A, the proposed condition of hidability is a stricter condition than (strong) information-theoretic secrecy. Furthermore, we also show (in appendix B) that the deniability and the hidability are independent of each other (neither one implies the other).

The protocols falling under the notion of *network anonymity*, such as Tor networks [3], enable users to route packets through crowds so that it is hard for eavesdroppers to estimate the source or the destination of the traffic. However, most anonymizing protocols (see *e.g.*, [4]) are not useful if Willie is eavesdropping at Alice's very point of connection to the network. For deniability, the active covert communication should look like an innocent behavior, and the very fact that Alice chooses to route packets through Tor might arouse Willie's suspicion.

Another notion of covert communication is *Steganography*, which involves hiding messages into the transmitted data such as images. The ideas presented in this paper come close to the information-theoretic steganography framework proposed in [5], [6], [7]. However, the main difference is that these steganographic protocols require (large) private keys to be shared between Alice and Bob, unlike our model, which requires no shared secret. But, contrary to our model which imposes restrictions on Willie's observation power, these steganographic protocols allow the eavesdropper to observe the entire network.

Our work falls in the line of [8] and [9]. In [8], the authors show that deniability can be achieved (under AWGN channel model) if Alice can secretly share her codebook with Bob. Che *et al.* [9] show that any shared secret is not necessary (under binary symmetric channel model), if the channel between Alice and Willie is noisier than that between Alice and Bob. Even though we use some information-theoretic techniques from [9] in this work, there are several key differences. First, in the model considered by [9] (and also [8]), in the innocent state, Alice opts to keep quiet (transmits the all zero codeword). On the other hand, in our work, we utilize Alice's innocent communication patterns to hide the covert messages. Secondly, in [9], the authors exploit the fact that Willie's channel is noisier than Bob's to achieve deniability (thus, hiding the covert data in noise), we exploit Willie's limited observing power to be deniable (thus, hiding the covert data in different channels). Finally, these papers only focus on deniability and do not consider the hidability metric.

The paper that comes closest to our work is [10]. We compare our results with those of [10] in the following.

- The notion of *effective secrecy* proposed by Hou and Kramer subsumes the notion of strong information-theoretic secrecy. Thus, their achievability scheme makes the system strongly information-theoretically secure. On the other hand, the hidability metric considered in our paper can be shown to be a stricter requirement than strong information-theoretic secrecy (see appendix A).
- The effective secrecy metric essentially combines together the metrics of deniability (referred to as *stealth*) and hidability (in particular strong secrecy), and thus, the encoding strategies considered in [10] always attain the deniability and the secrecy together. On the contrary, our encoding techniques allow us to maintain separability between the deniability and the hidability – it is possible to achieve the deniability without achieving the hidability and vice-versa.²
- Hou and Kramer consider the generic wiretap channel model. Our model, in which Willie can observe an unknown subset of links, can be considered as the model of an arbitrarily varying channel (AVC), where the capacity of the channel is a function of the subset that Willie is tapping.
- While the capacity result of Hou and Kramer [10] is more generic from the perspective of the channel model, we specialize the capacity for deniable and hidable communication for the specific channel such that the capacity expression is more explicitly computable.

²In principle, Hou and Kramer [10] could have designed the separate encoding schemes for deniability and secrecy, but their choice of metric, which merges the two conditions, leads to the single coding scheme achieving the both.

III. PROBLEM FORMULATION

A. Notational Conventions

Boldface upper-case symbols, *e.g.* $\mathbf{X}$, denote random variables, boldface lower-case symbols, *e.g.* $\mathbf{x}$, denote particular realizations of those random variables. Calligraphic symbols like $\mathcal{W}$ denote sets. The size of a set $\mathcal{W}$ is denoted as $|\mathcal{W}|$ and the complement is denoted as $\mathcal{W}^c$. Boldface symbols with arrow on their top, *e.g.*, $\vec{\mathbf{x}}$, denote vectors. We also use boldface symbols with an arrow on top to denote binary matrices. The distinction as to whether such symbol represents a vector or a binary matrix will be made clear. The reason behind using the same notation for vectors and binary matrices is that, in the latter part of the paper, we consider $m \times n$ binary matrices as length- n vectors of symbols chosen from the finite field $GF(2^m)$. Unless otherwise specified, all vectors are of length n , where n corresponds to the *block-length* (number of network uses). The probabilities of the events are denoted by symbol $\Pr$ with a subscript denoting the random variable(s) over which the probabilities are calculated.

We use standard notation for information-theoretic quantities: $H(\cdot)$ denotes the entropy function, $H(\cdot|\cdot)$ denotes the conditional entropy function, $I(\cdot;\cdot)$ denotes the mutual information, and $D(\cdot||\cdot)$ denotes the Kullback-Leibler divergence between two distributions. Lastly, we use the notation $\mathbb{V}(p_1(\cdot), p_2(\cdot))$ to denote the total variation distance between any two distributions $p_1(\cdot)$ and $p_2(\cdot)$, defined over an alphabet $\mathcal{X}$, defined as follows:

$$\mathbb{V}(p_1(\cdot), p_2(\cdot)) = \frac{1}{2} \sum_{\mathbf{x} \in \mathcal{X}} |p_1(\mathbf{x}) - p_2(\mathbf{x})|. \quad (1)$$

B. Problem Statement

Suppose Alice wants to (potentially) communicate with Bob over a *parallel link* network consisting of C parallel links. Each link is assumed to have the capacity of one bit per use³ and is assumed to be noiseless⁴. Alice is allowed to transmit a length- n binary sequence $\vec{\mathbf{x}}_i \in \{0, 1\}^{1 \times n}$ on the i -th link (here, i is an index in $\{1, \dots, C\}$) over n network uses (here, n is the *block-length*). Bob receives the set of C sequences and organizes them as the $C \times n$ binary received *codeword matrix* as $\vec{\mathbf{X}} = [\vec{\mathbf{x}}_1^T \vec{\mathbf{x}}_2^T \dots \vec{\mathbf{x}}_C^T]^T$.

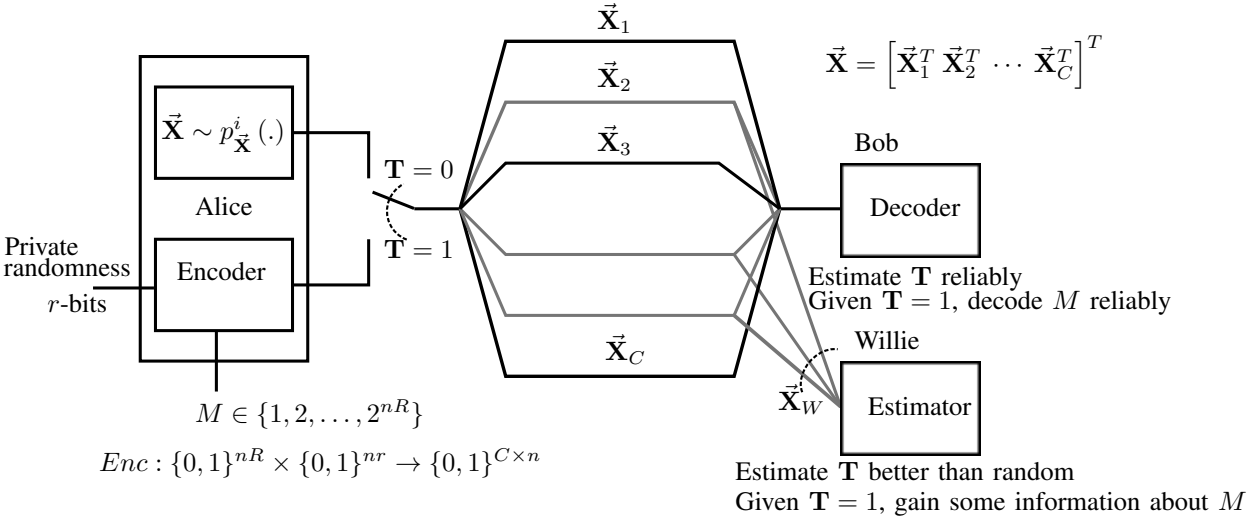


Fig. 1. System Model

Alice may or may not wish to communicate covertly with Bob, and accordingly she is said to be in an *active* or in an *innocent* state. If Alice does not have any covert message to transmit to Bob, she is said

³Note that, if the links have unequal capacities, it is possible to split each link into multiple links with the same capacity.

⁴It is worth noting that i.i.d. noise on links does not fundamentally weaken our results. In this case, one can use link-by-link error correction. However, in this paper, we do not explore along this direction.

to be in an innocent state. In the innocent state, during the t -th time instant ($1 \leq t \leq n$), Alice randomly generates a length- C binary *codeword* $\mathbf{x}(t) = [\mathbf{x}_1(t) \ \mathbf{x}_2(t) \ \cdots \ \mathbf{x}_C(t)]^T$, denoting the transmissions on each of the links, according to the distribution $p_{\mathbf{x}(t)}^i(\mathbf{x}(t))$ (over an alphabet of size 2^C) called as the *scalar innocent distribution*. We assume that the codewords $\mathbf{x}(t)$ are independent and identically distributed (i.i.d.) over the time instants t , $1 \leq t \leq n$.⁵ Therefore, the $C \times n$ binary codeword matrix $\vec{\mathbf{x}}$ that is transmitted by Alice is distributed according to the distribution $p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}}) = p_{\mathbf{x}}^i(\mathbf{x}(1)) p_{\mathbf{x}}^i(\mathbf{x}(2)) \cdots p_{\mathbf{x}}^i(\mathbf{x}(n))$, wherein $\vec{\mathbf{x}} = [\mathbf{x}(1) \ \mathbf{x}(2) \ \cdots \ \mathbf{x}(n)]$. This distribution $p_{\vec{\mathbf{x}}}^i(\cdot)$, defined over the alphabet of all binary $C \times n$ binary matrices, is called as the *innocent distribution*.

When Alice is in the active state, she wants to transmit a *covert message* $M \in \{1, \dots, 2^{nR}\}$ to Bob. We assume that the message symbols are distributed uniformly. Let $\vec{\mathbf{M}}$ be the random variable corresponding to the length- nR binary vector representing the covert message – here, R denotes the *rate* of Alice's covert message. Alice encodes her covert messages using an *encoder* $Enc : \{0, 1\}^{nR} \times \{0, 1\}^{nr} \rightarrow \{0, 1\}^{C \times n}$, where r denotes the *rate of private randomness* used by her encoder. The set $\{\vec{\mathbf{x}}_1, \dots, \vec{\mathbf{x}}_{|\mathcal{C}|}\}$ of all the possible output codeword matrices of the encoder Enc forms the *codebook* $\mathcal{C}$ for Alice. Notice that Alice's encoder induces a distribution on the transmitted binary codeword matrices. This distribution $\hat{p}_{\vec{\mathbf{x}}}(\cdot)$, defined over the alphabet of all binary $C \times n$ matrices, that is induced by Alice's encoding process in the active state is called as the *induced distribution*.

We use a binary random variable $\mathbf{T}$ to denote Alice's transmission status, with $\mathbf{T} = 0$ means Alice is innocent and $\mathbf{T} = 1$ means she is active. We assume that the prior statistics on $\mathbf{T}$ can be known to Willie, but need not be known to Bob. Further, we assume that only Alice knows the value of $\mathbf{T}$ *a priori*.

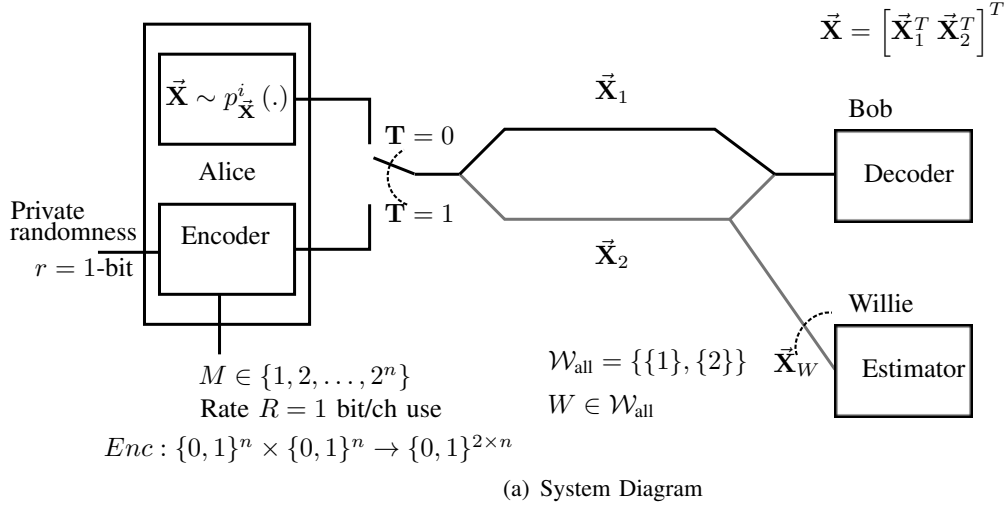
Communication takes place in the presence of a passive eavesdropper Willie, who can observe some subset of links. Let $\mathcal{W}$ denote the set of links that are eavesdropped by Willie, and $\mathcal{W}_{\text{all}}$ denote the class of all possible subsets of links which Willie can observe. For example, $\mathcal{W}_{\text{all}}$ might comprise of all subsets of at most $C - 1$ links. Let $\vec{\mathbf{x}}_{\mathcal{W}} \in \{0, 1\}^{|\mathcal{W}| \times n}$ be the codeword (sub-)matrix that is observed by Willie. The marginal distribution on the codewords that Willie observes when Alice is innocent, denoted as $p_{\vec{\mathbf{x}}_{\mathcal{W}}}^i(\cdot)$, defined over the alphabet of all the binary $|\mathcal{W}| \times n$ matrices, is called as the *marginal innocent distribution*. Similarly, the marginal distribution on the codewords observed by Willie when Alice is active, denoted as $\hat{p}_{\vec{\mathbf{x}}_{\mathcal{W}}}(\cdot)$, defined over the alphabet of all binary $|\mathcal{W}| \times n$ matrices, is called as the *marginal induced distribution*.

Throughout, we assume that there is *no* shared secret between Alice and Bob that is not known to Willie. Further, we assume that Willie knows the encoding-decoding scheme used by Alice and Bob.

Toy Example: Consider a parallel-link network consisting of $C = 2$ links as shown in Fig. 2. Suppose that Willie can eavesdrop on any one of the links. Thus, $\mathcal{W}_{\text{all}} = \{\{1\}, \{2\}\}$. When innocent, at each time instant t , $1 \leq t \leq n$, Alice choses length-2 binary codewords according to the scalar innocent distribution given in Fig. 2(b). Note that the innocent distribution on binary $2 \times n$ codeword matrices will be the product distribution given by $p_{\vec{\mathbf{x}}}^i(\cdot) = \{p_{\mathbf{x}}^i\}^n$. The marginal innocent distribution on the top link will be the Bernoulli process with parameter $\frac{1}{2}$, while the marginal innocent distribution on the bottom link will be the Bernoulli process with parameter $\frac{3}{4}$.

Suppose Alice wants to transmit a binary, length- n covert message $\vec{\mathbf{M}} \in \{0, 1\}^n$ to Bob. Notice that this results in Alice's transmission rate to be $R = 1$. Suppose that she uses Shannon's one-time padding scheme as her stochastic encoder, described as follows. For each time instant t , Alice generates a uniform random bit $K(t)$. On the top link, she transmits the random bit $K(t)$, whereas, on the bottom link, she transmits exclusive OR of the covert message bit and the random bit $M(t) \oplus K(t)$. With this encoding, the induced distribution $\hat{p}_{\vec{\mathbf{x}}}(\cdot)$ will be the uniform distribution over all the binary $2 \times n$ matrices. The marginal induced distribution on each link will also be the uniform distribution over the length- n binary vectors.

⁵One can question this assumption that the binary vectors transmitted over the parallel-link network are i.i.d. over the time instants t , $1 \leq t \leq n$. However, we should point out that getting the results under this (i.i.d. assumption) model is still challenging; we aim to extend this model to more realistic source models in the future work.



$\mathbf{X}(t) = [\mathbf{X}_1(t) \ \mathbf{X}_2(t)]^T$	$p_{\mathbf{X}}^i(\cdot)$
0 0	0
0 1	$\frac{1}{2}$
1 0	$\frac{1}{4}$
1 1	$\frac{1}{4}$

(b) Scalar Innocent Distribution

$\mathbf{X}(t) = [\mathbf{X}_1(t) \ \mathbf{X}_2(t)]^T$	$\hat{p}_{\mathbf{X}}(\cdot)$
0 0	$\frac{1}{4}$
0 1	$\frac{1}{4}$
1 0	$\frac{1}{4}$
1 1	$\frac{1}{4}$

(c) Scalar Induced Distribution

Fig. 2. Toy Example of a parallel-link network with $C = 2$ links. Willie can observe any one of the links.

We aim to design a communication scheme that satisfies the following three requirements:

1) *Reliability* : Bob should be able to estimate whether Alice is active or innocent and decode the messages correctly when Alice is active, with high probability over Alice's encoding scheme. This is, in fact, a basic requirement for any reasonable communication scheme, but we make it explicit here, and formally define this below.

Definition 1 (Reliability). *We say that the scheme is $(1 - \epsilon_r)$ -reliable if, for any arbitrarily small $\epsilon_r > 0$,*

$$\Pr_{\vec{\mathbf{X}}} \left(\hat{\mathbf{T}}_B = 1 | \mathbf{T} = 0 \right) + \Pr_{\vec{\mathbf{X}}} \left(\hat{\mathbf{T}}_B = 0 | \mathbf{T} = 1 \right) + \Pr_{\vec{\mathbf{M}}, \text{Enc}} \left(\hat{\mathbf{M}}_B \neq \vec{\mathbf{M}} | \mathbf{T} = 1 \right) \leq \epsilon_r, \quad (2)$$

where $\hat{\mathbf{T}}_B$ and $\hat{\mathbf{M}}_B$ denote Bob's estimates of Alice's transmission status $\mathbf{T}$ and her covert message $\vec{\mathbf{M}}$, respectively.

The first term in the summation in (2) gives the probability, over the randomness of the codeword matrices, that Bob estimates Alice to be active conditioned on the fact that she, actually, is innocent; the second term is the probability, over the randomness of the codeword matrices, that Bob finds Alice to be innocent given that she is indeed active; and the last term is the conditional probability, over the randomness of the covert messages and the private randomness of Alice's encoder, given that Alice is active, that Bob wrongly decodes the covert message. Notice that each of these three events is an error event, and for the scheme to be reliable, the probability of each of the error events should be very small, as defined above.

2) *Deniability* : Willie should not be able to "reliably" estimate Alice's transmission status $\mathbf{T}$, as formally defined below.

Definition 2 (Deniability). *We say that the scheme is $(1 - \epsilon_d)$ -deniable if, for any arbitrarily small $\epsilon_d > 0$,*

$$\mathbb{V} \left(p_{\vec{\mathbf{X}}_W}^i(\cdot), \hat{p}_{\vec{\mathbf{X}}_W}(\cdot) \right) < \epsilon_d, \quad \forall \mathcal{W} \in \mathcal{W}_{\text{all}}, \quad (3)$$

where $\mathbb{V}(\cdot, \cdot)$ denotes the total variation distance defined in (1).

$\mathbf{T}$	Alice's transmission status
$\hat{\mathbf{T}}_B$	Bob's estimate of Alice's transmission status
$\hat{\mathbf{T}}_W$	Willie's estimate of Alice's transmission status
C	Number of parallel links in the network
$\mathcal{W}_{\text{all}}$	Class of all possible subsets of links which Willie can observe
$\mathcal{W}$	Subset of links observed by Willie
M	Number of messages Alice wishes to transmit when active
$\vec{\mathbf{m}}$	Particular message transmitted by Alice (n-bit long)
$\vec{\mathbf{M}}$	Random variable (as a binary vector) corresponding to message
$\hat{\mathbf{M}}_B$	Bob's estimate of the transmitted message
$\hat{\mathbf{M}}_W$	Willie's estimate of the transmitted message
$\vec{\mathbf{x}}$	Codeword transmitted by Alice
$\vec{\mathbf{X}}$	Random variable corresponding to codeword
$\mathcal{X}$	Alphabet for codeword symbol
$p_{\vec{\mathbf{X}}}^i(\mathbf{x})$	Innocent distribution on codeword symbols
$p_{\vec{\mathbf{X}}}^i(\vec{\mathbf{x}})$	Innocent distribution on n-symbol length codeword
$\hat{p}_{\vec{\mathbf{X}}}(\vec{\mathbf{x}})$	Distribution induced by Alice's encoding in active state
$\vec{\mathbf{x}}_{\mathcal{W}}$	Codeword observed by Willie
$\vec{\mathbf{X}}_{\mathcal{W}}$	Random variable corresponding to the codeword observed by Willie
$\mathcal{X}_{\mathcal{W}}$	Alphabet for codeword symbols observed by Willie
$p_{\vec{\mathbf{X}}_{\mathcal{W}}}^i(\vec{\mathbf{x}}_{\mathcal{W}})$	Marginal innocent distribution observed by Willie
$\hat{p}_{\vec{\mathbf{X}}_{\mathcal{W}}}(\vec{\mathbf{x}}_{\mathcal{W}})$	Marginal active distribution observed by Willie
$\vec{\mathbf{x}}_{\mathcal{W}^c}$	Codeword that Willie cannot observe
$\mathcal{X}_{\mathcal{W}^c}$	Alphabet for codeword symbols that Willie cannot observe
$\mathcal{T}_{\epsilon}^{(n)}$	Strongly typical set of codewords
Q	Type of a codeword and also type class corresponding to that type
$\mathcal{T}_{\epsilon}^{(n)}(\vec{\mathbf{X}}_{\mathcal{W}^c} \vec{\mathbf{x}}_{\mathcal{W}})$	Conditionally strongly typical set

TABLE I
NOTATION USED THROUGHOUT THE PAPER

To get some intuition behind the definition of deniability, notice that Willie essentially performs binary hypothesis testing on the parameter $\mathbf{T}$. One can show, by standard hypothesis testing arguments [11], that if $\mathbb{V}\left(p_{\vec{\mathbf{X}}_{\mathcal{W}}}^i(\cdot), \hat{p}_{\vec{\mathbf{X}}_{\mathcal{W}}}(\cdot)\right)$ is upper bounded by some small ϵ_d , then Willie's best estimator based on his observations is at most ϵ_d better than even a naïve estimator independent of his network observations.

3) *Hidability* : When Alice is active, Willie's observations on the links in $\mathcal{W}$ should, with high probability, not “leak any information” about the message Alice is transmitting, as defined in Definition 3 below.

Definition 3 (Hidability). *We say that the scheme is $(1 - \epsilon_h)$ -hidable if, for any arbitrarily small ϵ_h , $0 < \epsilon_h < 1$, we have*

$$1 - \epsilon_h \leq \frac{\Pr_{\vec{\mathbf{M}}, Enc}(\vec{\mathbf{M}} = \vec{\mathbf{m}}|\vec{\mathbf{x}}_{\mathcal{W}}, \mathbf{T} = 1)}{\Pr_{\vec{\mathbf{M}}}(\vec{\mathbf{M}} = \vec{\mathbf{m}}|\mathbf{T} = 1)} \leq 1 + \epsilon_h, \quad \forall \vec{\mathbf{m}}, \forall \mathcal{W} \in \mathcal{W}_{\text{all}}, \quad (4)$$

for any binary $|\mathcal{W}| \times n$ matrix $\vec{\mathbf{x}}_{\mathcal{W}}$ that Willie observes.

Note that our definition of hidability is somewhat stronger than the usual definition of information-theoretic secrecy, since, as we show in Appendix A, $(1 - o(\frac{1}{n}))$ -hidability implies information-theoretic secrecy, but the converse is not necessarily true. Furthermore, at first sight, it seems like the deniability is a stronger condition than the hidability because deniability requires that the transmission status $\mathbf{T}$ should be indistinguishable to Willie irrespective of whether he can decode the codewords or not, but actually neither implies the other (see Appendix B for examples of hidable schemes that are not deniable, and deniable schemes that are not hidable.)

IV. DENIABLE ENCODERS

In this section, we focus on designing the schemes that achieve only deniability. In further sections, we extend these schemes to be hidable as well as deniable.

A. Main Result for Deniable Communication

We assume that the scalar induced distribution $p_{\mathbf{X}}^i(\cdot)$, defined over a finite alphabet $\mathcal{X} = \{0, 1, \dots, 2^C - 1\}$, is given. Then, the innocent distribution $p_{\vec{\mathbf{X}}}^i(\cdot)$, which is the product distribution $p_{\vec{\mathbf{X}}}^i(\cdot) = \prod_{t=1}^n p_{\mathbf{X}}^i(\cdot)$ (under the assumption that the codewords are i.i.d. over time instants), also gets fixed. We want to know at what rate Alice can transmit the covert messages and how she can perform encoding such that her active status is deniable from Bob.

Before characterizing the achievable deniable rate, we need to review several information-theoretic concepts related to *typicality* (see [12] for details). Hereafter, we consider the C parallel links as one hyperlink with input (output) alphabet as $\mathcal{X} = \{0, 1, \dots, 2^C - 1\}$. Thus, all the codewords will be length- n vectors with each symbol belonging to alphabet $\mathcal{X}$. Let us begin with the notion of type. The *type* Q of a sequence $\vec{\mathbf{x}} \in \mathcal{X}^n$ is the empirical distribution on $\mathcal{X}$ defined as [12]:

$$Q_{\vec{\mathbf{x}}} = \frac{1}{n} N(a|\vec{\mathbf{x}}) \quad \forall a \in \mathcal{X}, \quad (5)$$

where $N(a|\vec{\mathbf{x}})$ denotes the number of occurrences of a symbol $a \in \mathcal{X}$ in the sequence $\vec{\mathbf{x}}$. The set of all sequences of type Q is called as *type class* Q . We use the same notation to denote the type and the type class corresponding to that type; the distinction will be clear from the context. For some small $\epsilon > 0$, the ϵ -strongly typical set (w.r.t. a distribution $p_{\mathbf{X}}(\cdot)$) is defined as follows [12]:

$$\mathcal{T}_{\epsilon}^{(n)}(p_{\mathbf{X}}(\cdot)) = \left\{ \vec{\mathbf{x}} : \begin{array}{ll} \left| \frac{1}{n} N(a|\vec{\mathbf{x}}) - p_{\mathbf{X}}(a) \right| \leq \frac{\epsilon}{|\mathcal{X}|} & \forall a \in \mathcal{X}, \text{ if } p_{\mathbf{X}}(a) > 0 \\ N(a|\vec{\mathbf{x}}) = 0, & \text{if } p_{\mathbf{X}}(a) = 0 \end{array} \right\}. \quad (6)$$

Note that the typical set can be viewed as a collection of the type classes, which satisfy the aforementioned constraint. It is worth noting that the probability of each sequence in a particular type class is the same.

In the following we present the matching inner and outer bounds on the rate of covert transmission for deniable communication. In the following theorem, we quantify the deniability of the stochastic encoder for multiple links.

Theorem 1. *Given any scalar innocent distribution $p_{\mathbf{X}}^i(\cdot)$, and a class of subsets $\mathcal{W}_{all}$ that can be eavesdropped, it is possible to make the communication $(1 - \epsilon_d)$ -deniable, for some arbitrarily small $0 < \epsilon_d < 1$, with probability at least $1 - 2^{-c_1 2^{\Theta n}}$, where $c_1 > 0$ is some small constant, provided the covert communication rate is upper bounded as*

$$R \leq \sup_{\substack{p_{\mathbf{X}}(\cdot): \forall \mathcal{W} \in \mathcal{W}_{all} \\ p_{\mathbf{X}_{\mathcal{W}}}(\cdot) = p_{\mathbf{X}_{\mathcal{W}}}^i(\cdot)}} H(p_{\mathbf{X}}(\cdot)). \quad (7)$$

B. Achievability

First, we present an achievability scheme. Let the distribution $p_{\mathbf{X}}^a(\cdot)$, defined on the alphabet $\mathcal{X} = \{0, 1, \dots, 2^C - 1\}$, denote the *scalar active distribution* that is used by Alice to generate her codeword symbols. Suppose Alice computes the scalar active distribution by solving the following convex optimization problem.

$$p_{\mathbf{X}}^a(\cdot) = \underset{\substack{p_{\mathbf{X}}(\cdot): \forall \mathcal{W} \in \mathcal{W}_{all} \\ p_{\mathbf{X}_{\mathcal{W}}}(\cdot) = p_{\mathbf{X}_{\mathcal{W}}}^i(\cdot)}}{\operatorname{argmax}} H(p_{\mathbf{X}}(\cdot)). \quad (8)$$

Note that the codeword sequences will be distributed according to the *active distribution* $p_{\vec{\mathbf{X}}}^a(\cdot) = \prod_{t=1}^n p_{\mathbf{X}}^a(\cdot)$, defined over the alphabet $\mathcal{X}^n$. Alice performs the encoding stochastically according to the following procedure.

Procedure 1 (Stochastic Encoder for Multiple Link). *Random Binning Based Scheme*

- 1) *Generate the codebook $\mathcal{C}$ as the set of all strongly typical codewords (w.r.t. scalar active distribution $p_{\mathbf{X}}^a(\cdot)$), defined as follows*

$$\mathcal{T}_{\epsilon}^{(n)}(p_{\mathbf{X}}^a(\cdot)) = \left\{ \vec{\mathbf{x}} : \begin{array}{ll} \left| \frac{1}{n} N(a|\vec{\mathbf{x}}) - p_{\mathbf{X}}^a(a) \right| \leq \frac{\epsilon}{|\mathcal{X}|} & \forall a \in \mathcal{X}, \text{ if } p_{\mathbf{X}}^a(a) > 0 \\ N(a|\vec{\mathbf{x}}) = 0, & \text{if } p_{\mathbf{X}}^a(a) = 0 \end{array} \right\}, \quad (9)$$

for some small $\epsilon > 0$.

- 2) *Fix the number of messages as $|\vec{\mathbf{M}}| = 2^{nR}$, where R denotes the transmission rate of the scheme.*
- 3) *Assign each codeword $\vec{\mathbf{x}} \in \mathcal{C}$ to one of 2^{nR} message bins uniformly at random.*
- 4) *Let $\mathcal{B}(\vec{\mathbf{m}})$ denote the set of codewords that have fallen in the bin corresponding to the message $\vec{\mathbf{m}}$. For a particular message $\vec{\mathbf{m}}$, stochastically choose a codeword from the bin $\mathcal{B}(\vec{\mathbf{m}})$ according to the following conditional distribution:*

$$p_{\vec{\mathbf{x}}|\vec{\mathbf{M}}}^{\text{stoch}}(\vec{\mathbf{x}}|\vec{\mathbf{m}}) = \begin{cases} \frac{p_{\vec{\mathbf{x}}}^a(\vec{\mathbf{x}})}{\sum_{\vec{\mathbf{x}}:\vec{\mathbf{x}} \in \mathcal{C} \cap \mathcal{B}(\vec{\mathbf{m}})} p_{\vec{\mathbf{x}}}^a(\vec{\mathbf{x}})} & \text{if } \vec{\mathbf{x}} \in \mathcal{B}(\vec{\mathbf{m}}), \\ 0 & \text{Otherwise.} \end{cases} \quad (10)$$

Note that the encoder $\text{Enc} : \vec{\mathbf{M}} \rightarrow \mathcal{C}$ is a one-to-many stochastic encoder defined by the aforementioned distribution $p_{\vec{\mathbf{x}}|\vec{\mathbf{M}}}^{\text{stoch}}(\cdot|\cdot)$.

Let us characterize the distribution $\hat{p}_{\vec{\mathbf{x}}}(\cdot)$ induced by the proposed encoding.

$$\hat{p}_{\vec{\mathbf{x}}}(\vec{\mathbf{x}}) \stackrel{(a)}{=} \sum_{\vec{\mathbf{m}}} \hat{p}_{\vec{\mathbf{x}}}(\vec{\mathbf{x}}, \vec{\mathbf{m}}), \quad (11)$$

$$= \sum_{\vec{\mathbf{m}}} p_{\vec{\mathbf{x}}|\vec{\mathbf{M}}}^{\text{stoch}}(\vec{\mathbf{x}}|\vec{\mathbf{m}}) \Pr_{\vec{\mathbf{M}}, \text{Enc}}(\vec{\mathbf{m}}), \quad (12)$$

$$\stackrel{(b)}{=} \frac{p_{\vec{\mathbf{x}}}^a(\vec{\mathbf{x}})}{\sum_{\vec{\mathbf{x}}:\vec{\mathbf{x}} \in \mathcal{C} \cap \mathcal{B}(\vec{\mathbf{m}})} p_{\vec{\mathbf{x}}}^a(\vec{\mathbf{x}})} \Pr_{\vec{\mathbf{M}}, \text{Enc}}(\vec{\mathbf{m}}_{\vec{\mathbf{x}}}), \quad (13)$$

$$\stackrel{(c)}{=} \frac{p_{\vec{\mathbf{x}}}^a(\vec{\mathbf{x}})}{\sum_{\vec{\mathbf{x}}:\vec{\mathbf{x}} \in \mathcal{C} \cap \mathcal{B}(\vec{\mathbf{m}})} p_{\vec{\mathbf{x}}}^a(\vec{\mathbf{x}})} \left(\frac{1}{|\vec{\mathbf{M}}|} \right) \quad (14)$$

where (a) follows from the law of total probability, (b) follows from the definition of the stochastic encoder (10), and (c) follows because the covert messages are uniformly distributed $\Pr(\vec{\mathbf{m}}) = \frac{1}{|\vec{\mathbf{M}}|} \forall \vec{\mathbf{m}}$.

Now, let us define the term $\Pr_{\mathcal{B}}(\mathcal{B}(\vec{\mathbf{m}})) = \sum_{\vec{\mathbf{x}}:\vec{\mathbf{x}} \in \mathcal{C} \cap \mathcal{B}(\vec{\mathbf{m}})} p_{\vec{\mathbf{x}}}^a(\vec{\mathbf{x}})$ as the probability of a particular message bin. Intuitively, the probability of a message bin is the sum of probabilities (under active distribution) of all the codewords that have fallen in that particular bin. From lemma 1 in appendix D, with

probability at least $1 - 2 \exp \left(\frac{-\epsilon^2 2^n \left[H(Q) - \frac{2^C \log_2(n+1)}{n} \right]}{3|\vec{\mathbf{M}}|} \right)$, the probability of a message bin can be bounded as

$$(1 - \epsilon) (1 - 2^{C+1} \exp(-2\epsilon^2 n)) \frac{1}{|\vec{\mathbf{M}}|} \leq \Pr_{\mathcal{B}}(\mathcal{B}(\vec{\mathbf{m}})) \leq (1 + \epsilon) \frac{1}{|\vec{\mathbf{M}}|}. \quad (15)$$

Since $\hat{p}_{\vec{\mathbf{x}}}(\vec{\mathbf{x}}) = \frac{p_{\vec{\mathbf{x}}}^a(\vec{\mathbf{x}})}{\Pr_{\mathcal{B}}(\mathcal{B}(\vec{\mathbf{m}}))|\vec{\mathbf{M}}|}$ (see (14)), we have, $\forall \vec{\mathbf{x}} \in \mathcal{T}_{\epsilon}^{(n)}(p_{\mathbf{X}}^a(\cdot))$,

$$\frac{p_{\vec{\mathbf{x}}}^a(\vec{\mathbf{x}})}{1 + \epsilon} \leq \hat{p}_{\vec{\mathbf{x}}}(\vec{\mathbf{x}}) \leq \frac{p_{\vec{\mathbf{x}}}^a(\vec{\mathbf{x}})}{(1 - \epsilon) (1 - 2^{C+1} \exp(-2\epsilon^2 n))} \text{ w.p. at least } 1 - 2 \exp \left(\frac{-\epsilon^2 2^n \left[H(Q) - \frac{2^C \log_2(n+1)}{n} \right]}{3|\vec{\mathbf{M}}|} \right). \quad (16)$$

Now, consider the total variation distance between the active and the induced distribution as follows:

$$\mathbb{V}(p_{\vec{\mathbf{x}}}^a(\cdot), \hat{p}_{\vec{\mathbf{x}}}(\cdot)) \stackrel{(d)}{=} \frac{1}{2} \sum_{\vec{\mathbf{x}}} |p_{\vec{\mathbf{x}}}^a(\vec{\mathbf{x}}) - \hat{p}_{\vec{\mathbf{x}}}(\vec{\mathbf{x}})| \quad (17)$$

$$= \frac{1}{2} \sum_{\vec{\mathbf{x}}: \vec{\mathbf{x}} \notin \mathcal{T}_\epsilon^{(n)}} |p_{\vec{\mathbf{x}}}^a(\vec{\mathbf{x}}) - \hat{p}_{\vec{\mathbf{x}}}(\vec{\mathbf{x}})| + \frac{1}{2} \sum_{\vec{\mathbf{x}}: \vec{\mathbf{x}} \in \mathcal{T}_\epsilon^{(n)}} |p_{\vec{\mathbf{x}}}^a(\vec{\mathbf{x}}) - \hat{p}_{\vec{\mathbf{x}}}(\vec{\mathbf{x}})| \quad (18)$$

$$\stackrel{(e)}{\leq} \frac{2^{C+1} \exp(-2\epsilon^2 n)}{2} + \frac{1}{2} \sum_{\vec{\mathbf{x}}: \vec{\mathbf{x}} \in \mathcal{T}_\epsilon^{(n)}} |p_{\vec{\mathbf{x}}}^a(\vec{\mathbf{x}}) - \hat{p}_{\vec{\mathbf{x}}}(\vec{\mathbf{x}})| \quad (19)$$

$$\stackrel{(f)}{\leq} 2^C \exp(-2\epsilon^2 n) + \frac{1}{2} \sum_{\vec{\mathbf{x}}: \vec{\mathbf{x}} \in \mathcal{T}_\epsilon^{(n)}} \left| \frac{1}{(1-\epsilon)(1-2^{C+1} \exp(-2\epsilon^2 n))} - 1 \right| p_{\vec{\mathbf{x}}}^a(\vec{\mathbf{x}}) \text{ w.h.p.} \quad (20)$$

$$\stackrel{(g)}{\leq} 2^C \exp(-2\epsilon^2 n) + \frac{1}{2} \left| \frac{1}{(1-\epsilon)(1-2^{C+1} \exp(-2\epsilon^2 n))} - 1 \right| \text{ w.h.p.} \quad (21)$$

$$\stackrel{(h)}{=} \frac{1 - (1-\epsilon)(1-2^{C+1} \exp(-2\epsilon^2 n))^2}{2(1-\epsilon)(1-2^{C+1} \exp(-2\epsilon^2 n))}, \quad (22)$$

with probability at least

$$1 - 2 \exp \left(\frac{-\epsilon^2 2^n \left[H(Q) - \frac{2^C \log_2(n+1)}{n} \right]}{3 |\vec{\mathbf{M}}|} \right). \quad (23)$$

Here, (d) follows from the definition of total variation distance (see (1)). For inequality (e), note that $\hat{p}_{\vec{\mathbf{x}}}(\vec{\mathbf{x}}) = 0 \forall \vec{\mathbf{x}} \notin \mathcal{T}_\epsilon^{(n)}$ since the proposed encoder never transmits any non-typical codeword. Thus, the first summation in (18) is essentially $\Pr_{\vec{\mathbf{x}}}(\vec{\mathbf{x}} \notin \mathcal{T}_\epsilon^{(n)})$, which is upper bounded by $2|\mathcal{X}| \exp(-2\epsilon^2 n)$ (see [12]). Inequality (f) follows from the bounds on $\hat{p}_{\vec{\mathbf{x}}}(\vec{\mathbf{x}})$ derived in (16). Inequality (g) follows from the fact that the second summation in (20) contains the term $\sum_{\vec{\mathbf{x}}: \vec{\mathbf{x}} \in \mathcal{T}_\epsilon^{(n)}} p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}}) = \Pr_{\vec{\mathbf{x}}}(\mathcal{T}_\epsilon^{(n)})$ which is less than unity. Finally, in (h), we precisely mention the lower bound on the probability given in (16).

Now, suppose the number of cover messages are upper bounded as $|\vec{\mathbf{M}}| \leq 2^{n[H(p_{\vec{\mathbf{x}}}^a(\cdot)) - \tilde{r}_d]}$ for some small $\tilde{r}_d > 0$; here, the constant $\tilde{r}_d$ is the penalty in the transmission rate because of the deniability requirement. Also, let $Q^* = \arg \min_{Q \in \mathcal{T}_\epsilon^{(n)}(\mathbf{X})} H(Q)$ is the type class in the typical set such that the entropy of the corresponding type is the minimum. Then, the probability in (23) can be further lower bounded as

$$1 - 2 \exp \left(\frac{-\epsilon^2 2^n \left[H(Q^*) - \frac{2^C \log_2(n+1)}{n} - H(p_{\vec{\mathbf{x}}}^a(\cdot)) + \tilde{r}_d \right]}{3} \right). \quad (24)$$

Note that this probability given above will approach 1 in double exponential rate if the term $H(Q^*) - \frac{2^C \log_2(n+1)}{n} - H(p_{\vec{\mathbf{x}}}^a(\cdot)) + \tilde{r}_d$ is positive. Thus, our aim is to characterize the value of $\tilde{r}_d$ such that the term is positive. Let $\theta(Q^*)$ denote the total variation distance between type Q^* and $p_{\vec{\mathbf{x}}}^i(\cdot)$. Then, we have

$$\theta(Q^*) = \mathbb{V}(Q^*, p_{\vec{\mathbf{x}}}^a(\cdot)) = \frac{1}{2} \sum_{a \in \mathcal{X}} |Q^*(a) - p_{\vec{\mathbf{x}}}^a(a)|, \quad (25)$$

$$\stackrel{(l)}{\leq} \frac{1}{2} \sum_{a \in \mathcal{X}} \frac{\epsilon}{|\mathcal{X}|} = \frac{\epsilon}{2}, \quad (26)$$

where inequality(1) follows because the type Q belongs to ϵ -strongly typical set. Now, for any $\epsilon \leq \frac{1}{2}$, we will have $\theta(Q^*) \leq \frac{1}{4}$, which would allow us to use the following result from [12] (Lemma 2.7):

$$|H(Q^*) - H(p_{\mathbf{X}}^a(\cdot))| \leq -2\theta(Q^*) \log_2 \frac{2\theta(Q^*)}{|\mathcal{X}|}. \quad (27)$$

Moreover, since the RHS in (27) is an increasing function of $\theta(Q^*)$ for $0 \leq \theta(Q^*) \leq \frac{1}{2}$, we have $|H(Q^*) - H(p_{\mathbf{X}}^i(\cdot))| \leq -\epsilon \log_2 \frac{\epsilon}{|\mathcal{X}|}$. Therefore, we get $H(Q^*) - \frac{2^C \log_2(n+1)}{n} - H(p_{\mathbf{X}}^a(\cdot)) + \tilde{r}_d \geq \epsilon \log_2 \frac{\epsilon}{2^C} - \frac{2^C \log_2(n+1)}{n} + \tilde{r}_d$, which is positive provided

$$\tilde{r}_d > \epsilon \log_2 \frac{2^C}{\epsilon} + \frac{2^C \log_2(n+1)}{n}. \quad (28)$$

Therefore, we can see that the bound on $\mathbb{V}(p_{\mathbf{X}}^a(\cdot), \hat{p}_{\mathbf{X}}(\cdot))$ in (22) holds with probability at least $1 - 2^{-c_1 2^{\Theta(n)}}$ for some suitable $c_1 > 0$. Next, using this result, it is straightforward to show that, $\forall \mathcal{W} \in \mathcal{W}_{\text{all}}$, the total variation distance $\mathbb{V}(p_{\mathbf{X}_{\mathcal{W}}}^a(\cdot), \hat{p}_{\mathbf{X}_{\mathcal{W}}}(\cdot))$ between the marginal active distribution and the marginal induced distribution also satisfies the same bound with the same probability.

Finally, to prove the deniability, note that, since Alice computes the scalar active distribution using (8), we have, $\forall \mathcal{W} \in \mathcal{W}_{\text{all}}$, $p_{\mathbf{X}_{\mathcal{W}}}^a(\cdot) = p_{\mathbf{X}_{\mathcal{W}}}^i(\cdot)$, and thus, $p_{\mathbf{X}_{\mathcal{W}}}^a(\cdot) = p_{\mathbf{X}_{\mathcal{W}}}^i(\cdot)$. Therefore, we have

$$\mathbb{V}(p_{\mathbf{X}_{\mathcal{W}}}^a(\cdot), \hat{p}_{\mathbf{X}_{\mathcal{W}}}(\cdot)) \leq \frac{1 - (1 - \epsilon)(1 - 2^{C+1} \exp(-2\epsilon^2 n))^2}{2(1 - \epsilon)(1 - 2^{C+1} \exp(-2\epsilon^2 n))}, \quad (29)$$

with probability at least $1 - 2^{-c_1 2^{\Theta(n)}}$ for some suitable $c_1 > 0$, provided

$$|\vec{\mathbf{M}}| \leq 2^n [H(p_{\mathbf{X}}^a(\cdot)) - \tilde{r}_d], \quad \text{where} \quad \tilde{r}_d > \epsilon \log_2 \frac{2^C}{\epsilon} + \frac{2^C \log_2(n+1)}{n}. \quad (30)$$

It is easy to verify that if (30) is satisfied, then for sufficiently large n , the total variation distance between the marginal innocent distribution and the marginal induced distribution can be made arbitrarily small with probability double exponentially close to 1 by choosing $\epsilon = \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$, which completes the proof for deniability.

Remark 1. Note that here we did not consider how Bob can reliably estimate $\mathbf{T}$ and successfully decode the covert messages if $\mathbf{T} = 1$. We modify the aforementioned scheme for reliability in appendix E.

C. Converse

We use the standard information-theoretic arguments as follows.

$$nR \leq H(\vec{\mathbf{M}} | \mathbf{T} = 1), \quad (31)$$

$$= I(\vec{\mathbf{M}}; \vec{\mathbf{X}} | \mathbf{T} = 1) + H(\vec{\mathbf{M}} | \vec{\mathbf{X}}, \mathbf{T} = 1), \quad (32)$$

$$\stackrel{(a)}{=} I(\vec{\mathbf{M}}; \vec{\mathbf{X}} | \mathbf{T} = 1), \quad (33)$$

$$= H(\vec{\mathbf{X}} | \mathbf{T} = 1) - H(\vec{\mathbf{X}} | \vec{\mathbf{M}}, \mathbf{T} = 1), \quad (34)$$

$$\stackrel{(b)}{\leq} H(\vec{\mathbf{X}} | \mathbf{T} = 1), \quad (35)$$

$$\stackrel{(c)}{\leq} \sum_{j=1}^n H(\vec{\mathbf{X}}(j) | \mathbf{T} = 1), \quad (36)$$

$$\stackrel{(d)}{\leq} n \sup_{\substack{p_{\mathbf{X}}(\cdot): \forall \mathcal{W} \in \mathcal{W}_{\text{all}} \\ p_{\mathbf{X}_{\mathcal{W}}}(\cdot) = p_{\mathbf{X}_{\mathcal{W}}}^i(\cdot)}} H(p_{\mathbf{X}}(\cdot) | \mathbf{T} = 1), \quad (37)$$

where (a) follows since the links are assumed to be noiseless⁶, (b) is due to non-negativity of entropy, (c) is due to the independence bound on entropy, and (d) is due to the deniability condition of matching the marginals⁷.

V. DENIABLE AND HIDABLE ENCODERS

While deniability ensures that Willie cannot infer Alice's transmission status, it does not guarantee the hidability. Even if Willie is unable to estimate the transmission status, he can try to infer some information about the potential messages. At the same time, the hidability does not ensure deniability, since the notion of hidability focusses on securing the covert messages but not the fact that covert communication is under progress. We present some examples of the protocols that are deniable but not hidable, and vice-versa in the appendix B.

A. Main Result for Deniable and Hidable Communication

In the following, we show that the requirement of hidability in addition to deniability further reduces the rate at which Alice can communicate covert messages.

Theorem 2. *Given any scalar innocent distribution $p_{\mathbf{X}}^i(\cdot)$, and a class of subsets $\mathcal{W}_{all}$ that can be eavesdropped, it is possible to make the communication simultaneously $(1 - \epsilon_d)$ -deniable and $(1 - \epsilon_h)$ -hidable, for some arbitrarily small $0 < \epsilon_d, \epsilon_h < 1$, with probability at least $1 - 2^{-c_1 2^{\Theta_n}}$, where $c_1 > 0$ is some small constant, provided the covert communication rate is upper bounded as*

$$R_s \leq \sup_{\substack{p_{\mathbf{X}}(\cdot): \forall \mathcal{W} \in \mathcal{W}_{all} \\ p_{\mathbf{X}_{\mathcal{W}}}(\cdot) = p_{\mathbf{X}_{\mathcal{W}}}^i(\cdot)}} \min_{\mathcal{W} \in \mathcal{W}_{all}} [H(p_{\mathbf{X}}(\cdot)) - H(p_{\mathbf{X}_{\mathcal{W}}}(\cdot))]. \quad (38)$$

B. Achievability

In the following, we show that the random binning based stochastic encoding scheme (given by procedure 1) can also be used to achieve hidability in addition to deniability, provided the rate of covert messages is reduced below a threshold (as quantified in theorem 2). We repeat the procedure for the sake of clarity of exposition.

Procedure 2 (Deniable-Hidable Stochastic Encoder). Random Binning Based Scheme

- 1) Generate the codebook $\mathcal{C}$ as the set of all strongly typical codewords for some small $\epsilon > 0$ defined as follows (cf. (6))

$$\mathcal{C} = \mathcal{T}_{\epsilon}^{(n)}(\vec{\mathbf{X}}) = \left\{ \vec{\mathbf{x}} : \begin{array}{ll} \left| \frac{1}{n} N(a|\vec{\mathbf{x}}) - p_{\mathbf{X}}^a(a) \right| \leq \frac{\epsilon}{|\mathcal{X}|} & \forall a \in \mathcal{X}, \text{ if } p_{\mathbf{X}}^a(a) > 0 \\ N(a|\vec{\mathbf{x}}) = 0, & \text{if } p_{\mathbf{X}}^a(a) = 0 \end{array} \right\}$$

Note that each codeword $\vec{\mathbf{x}} \in \mathcal{X}^n$, where $\mathcal{X} = \{0, 1, \dots, 2^C - 1\}$.

- 2) Consider the set of secret messages as $\vec{\mathbf{M}} = \{1, \dots, 2^{nR_s}\}$. Associate every message with a bin, resulting in 2^{nR_s} number of bins.
- 3) Assign every codeword $\vec{\mathbf{x}} \in \mathcal{C}$ to one of 2^{nR_s} bins uniformly at random.
- 4) Let $\mathcal{B}(\vec{\mathbf{m}})$ denote the set of codewords that have fallen in the bin corresponding to the message $\vec{\mathbf{m}}$. In order to transmit a secret message $\vec{\mathbf{m}}$, stochastically choose a codeword from the bin $\mathcal{B}(\vec{\mathbf{m}})$ according to the following conditional distribution:

$$p_{\vec{\mathbf{X}}|\vec{\mathbf{M}}}^{stoch}(\vec{\mathbf{x}}|\vec{\mathbf{m}}) = \begin{cases} \frac{p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}})}{\sum_{\vec{\mathbf{x}}: \vec{\mathbf{x}} \in \mathcal{B}(\vec{\mathbf{m}})} p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}})} & \text{if } \vec{\mathbf{x}} \in \mathcal{B}(\vec{\mathbf{m}}), \\ 0 & \text{Otherwise.} \end{cases} \quad (39)$$

⁶In case there exist i.i.d. noise on each link, one can use Fano's inequality, and the rest of the arguments remain the same.

⁷In principle, the maximization in (37) can be carried out over all $p_{\mathbf{X}}(\cdot)$ such that $\mathbb{V}(p_{\mathbf{X}_{\mathcal{W}}}(\cdot), p_{\mathbf{X}_{\mathcal{W}}}^i(\cdot)) \leq \frac{\epsilon_d}{2} \forall \mathcal{W} \in \mathcal{W}_{all}$.

Note that the encoder $Enc : \vec{\mathbf{M}} \rightarrow \mathcal{C}$ is a one-to-many stochastic encoder defined by the aforementioned distribution $p_{\vec{\mathbf{X}}|\vec{\mathbf{M}}}^{stoch}(\cdot|\cdot)$.

We show that the scheme is reliable, deniable, and hidable in appendix C.

C. Converse

The converse follows from the standard information theoretic arguments as follows.

$$nR_s \leq H(\vec{\mathbf{M}}) \quad (40)$$

$$\stackrel{(a)}{\leq} H(\vec{\mathbf{M}}|\vec{\mathbf{X}}_{\mathcal{W}}) \quad (41)$$

$$= H(\vec{\mathbf{M}}|\vec{\mathbf{X}}_{\mathcal{W}}) - H(\vec{\mathbf{M}}|\vec{\mathbf{X}}_{\mathcal{W}}, \vec{\mathbf{X}}_{\mathcal{W}^c}) + H(\vec{\mathbf{M}}|\vec{\mathbf{X}}_{\mathcal{W}}, \vec{\mathbf{X}}_{\mathcal{W}^c}) \quad (42)$$

$$\stackrel{(b)}{=} H(\vec{\mathbf{M}}|\vec{\mathbf{X}}_{\mathcal{W}}) - H(\vec{\mathbf{M}}|\vec{\mathbf{X}}_{\mathcal{W}}, \vec{\mathbf{X}}_{\mathcal{W}^c}) \quad (43)$$

$$= I(\vec{\mathbf{M}}; \vec{\mathbf{X}}_{\mathcal{W}^c}|\vec{\mathbf{X}}_{\mathcal{W}}) \quad (44)$$

$$= H(\vec{\mathbf{X}}_{\mathcal{W}^c}|\vec{\mathbf{X}}_{\mathcal{W}}) - H(\vec{\mathbf{M}}|\vec{\mathbf{X}}_{\mathcal{W}}, \vec{\mathbf{M}}) \quad (45)$$

$$\stackrel{(c)}{\leq} H(\vec{\mathbf{X}}_{\mathcal{W}^c}|\vec{\mathbf{X}}_{\mathcal{W}}) \quad (46)$$

$$\stackrel{(d)}{\leq} n \min_{\mathcal{W} \in \mathcal{W}_{\text{all}}} H(\mathbf{X}_{\mathcal{W}^c}|\mathbf{X}_{\mathcal{W}}), \quad (47)$$

where (a) follows from the requirement of hidability (note that the hidability condition always guarantees the strong information-theoretic secrecy), (b) follows because the encoding is uniquely decodable (for noisy links one can use Fano's inequality), (c) follows from the non-negativity of entropy, and (d) is due to independence bound of the entropy. The minimization arises from Willie's ability to tap the subset of links having maximum entropy.

VI. CONCLUSION

In this paper, we proposed the concept of deniability for the parallel-link networks. We characterized the capacity of for the deniable communication over a parallel-link network and presented an achievability strategy based on random binning based stochastic encoding. Further, we proposed the concept of hidability, which is a stronger concept than the strong information-theoretic secrecy. We characterized the capacity for the (simultaneous) deniable and hidable communication. We also presented a scheme that allows the receiver to reliably identify the presence of the covert messages and to successfully decode the covert messages(if they are being transmitted) with asymptotically negligible penalty in the rate.

APPENDIX A

HIDABILITY VS. INFORMATION THEORETIC SECRECY

In this appendix, we compare the condition of hidability with the condition for strong information-theoretic secrecy. Recall that a scheme is said to be $(1 - \epsilon_h)$ -hidable if, for some small $\epsilon_h > 0$, we have

$$\frac{\Pr_{\vec{\mathbf{M}}, Enc}(\vec{\mathbf{M}} = \vec{\mathbf{m}}|\vec{\mathbf{X}}_{\mathcal{W}}, \mathbf{T} = 1)}{\Pr_{\vec{\mathbf{M}}}(\vec{\mathbf{M}} = \vec{\mathbf{m}}|\mathbf{T} = 1)} \in [1 - \epsilon_h, 1 + \epsilon_h], \quad \forall \vec{\mathbf{m}}, \forall \mathcal{W} \in \mathcal{W}_{\text{all}}, \quad (48)$$

for any $|\mathcal{W}| \times n$ binary matrix $\vec{\mathbf{x}}_{\mathcal{W}}$ that is observed by Willie. On the other hand, for any scheme to satisfy strong information-theoretic secrecy, we need to have [2]:

$$I\left(\vec{\mathbf{M}}; \vec{\mathbf{X}}_{\mathcal{W}} | \mathbf{T} = 1\right) \leq \epsilon_s, \quad \forall \mathcal{W} \in \mathcal{W}_{\text{all}}, \quad (49)$$

for some arbitrarily small $\epsilon_s > 0$.

In the following, we show that the hidability conditional is a stronger condition than the strong information-theoretic secrecy condition.

Theorem 3. *If a communication scheme is $(1 - o(\frac{1}{n}))$ -hidable, then it must have the strong information-theoretic security. The converse is not necessarily true.*

Proof: Suppose that the communication scheme under consideration is $(1 - \epsilon_h)$ -hidable. Consider the entropy of the covert messages conditioned on Willie's particular observation given that Alice is in the active status, as follows:

$$H\left(\vec{\mathbf{M}} | \vec{\mathbf{X}}_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}}, \mathbf{T} = 1\right) = - \sum_{\vec{\mathbf{m}}} \Pr_{\vec{\mathbf{M}}, \text{Enc}}(\vec{\mathbf{m}} | \vec{\mathbf{x}}_{\mathcal{W}}, \mathbf{T} = 1) \log_2 \left[\Pr_{\vec{\mathbf{M}}, \text{Enc}}(\vec{\mathbf{m}} | \vec{\mathbf{x}}_{\mathcal{W}}, \mathbf{T} = 1) \right] \quad (50)$$

$$\stackrel{(a)}{\geq} - \sum_{\vec{\mathbf{m}}} (1 - \epsilon_h) \Pr_{\vec{\mathbf{M}}}(\vec{\mathbf{m}} | \mathbf{T} = 1) \log_2 [(1 + \epsilon_h) \Pr_{\vec{\mathbf{M}}}(\vec{\mathbf{m}} | \mathbf{T} = 1)] \quad (51)$$

$$= (1 - \epsilon_h) H\left(\vec{\mathbf{M}} | \mathbf{T} = 1\right) - (1 - \epsilon_h) \log_2 (1 + \epsilon_h), \quad (52)$$

where (a) follows from the assumption that the scheme is $(1 - \epsilon_h)$ -hidable. Then, we have

$$H\left(\vec{\mathbf{M}} | \vec{\mathbf{X}}_{\mathcal{W}}, \mathbf{T} = 1\right) = \sum_{\vec{\mathbf{x}}_{\mathcal{W}}} \Pr_{\vec{\mathbf{M}}, \text{Enc}}(\vec{\mathbf{x}}_{\mathcal{W}} | \mathbf{T} = 1) H\left(\vec{\mathbf{M}} | \vec{\mathbf{X}}_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}}, \mathbf{T} = 1\right) \quad (53)$$

$$\stackrel{(b)}{\geq} \sum_{\vec{\mathbf{x}}_{\mathcal{W}}} \Pr_{\vec{\mathbf{M}}, \text{Enc}}(\vec{\mathbf{x}}_{\mathcal{W}} | \mathbf{T} = 1) (1 - \epsilon_h) \left[H\left(\vec{\mathbf{M}} | \mathbf{T} = 1\right) - \log_2 (1 + \epsilon_h) \right] \quad (54)$$

$$= (1 - \epsilon_h) \left[H\left(\vec{\mathbf{M}} | \mathbf{T} = 1\right) - \log_2 (1 + \epsilon_h) \right], \quad (55)$$

where (b) follows from the lower bound (52). Therefore, we can write

$$I\left(\vec{\mathbf{M}}; \vec{\mathbf{X}}_{\mathcal{W}} | \mathbf{T} = 1\right) = H\left(\vec{\mathbf{M}} | \mathbf{T} = 1\right) - H\left(\vec{\mathbf{M}} | \vec{\mathbf{X}}_{\mathcal{W}}, \mathbf{T} = 1\right) \quad (56)$$

$$\stackrel{(c)}{\leq} H\left(\vec{\mathbf{M}} | \mathbf{T} = 1\right) - (1 - \epsilon_h) \left[H\left(\vec{\mathbf{M}} | \mathbf{T} = 1\right) - \log_2 (1 + \epsilon_h) \right] \quad (57)$$

$$= \epsilon_h H\left(\vec{\mathbf{M}} | \mathbf{T} = 1\right) + (1 - \epsilon_h) \log_2 (1 + \epsilon_h) \quad (58)$$

$$\stackrel{(d)}{\leq} \epsilon_h n R_s + (1 - \epsilon_h) \log_2 (1 + \epsilon_h), \quad (59)$$

where (c) follows from (55), and (d) follows because $H\left(\vec{\mathbf{M}} | \mathbf{T} = 1\right) \leq \log_2 |\vec{\mathbf{M}}| = n R_s$, where R_s is the rate of transmission for the covert messages. Notice that any of the arguments above does not depend upon the specific choice of $\mathcal{W}$. Hence, $\forall \mathcal{W} \in \mathcal{W}_{\text{all}}$, we have $I\left(\vec{\mathbf{M}}; \vec{\mathbf{X}}_{\mathcal{W}} | \mathbf{T} = 1\right) \leq \epsilon_h n R_s + \mathcal{O}(\epsilon_h)$, and thus, if we choose $\epsilon_h = \frac{\epsilon_s}{n R_s}$, we will achieve information-theoretic secrecy. Moreover, if $\epsilon_h = o(\frac{1}{n})$, the hidability implies the strong information-theoretic security for any arbitrarily small ϵ_s .

To show that the converse may not necessarily be true, we give an example of a strongly secure scheme that is not hidable. Even though we do not worry about the deniability in this example, it is possible to extend the presented approach to deniable schemes. Consider a $C = 2$ link parallel-link network, with Willie observing any one of the links. In the active state, Alice employs a slight variant of Shannon's one-time padding scheme as follows.

Alice, first, randomly selects a subset $\mathcal{D}$ of binary, length- n sequences of cardinality $|\mathcal{D}| = 2^{n\delta}$ for some small $\delta \geq 0$. While transmitting a particular covert message sequence $\vec{m}$, if $\vec{m} \notin \mathcal{D}$, Alice uses Shannon's one-time padding. Specifically, she generates a binary, length- n key sequence $\vec{k}$ uniformly at random from the set $\mathcal{D}^c$. Then, she transmits the key sequence $\vec{k}$ on the top link, and the exclusive OR of the key sequence and the covert message on the bottom link. However, if $\vec{m} \in \mathcal{D}$, she simply transmits the message $\vec{m}$ on the top as well as the bottom link.

First, let us show that this scheme is strongly secure. Notice that, since Willie knows the encoding scheme, he knows the set $\mathcal{D}$. Without loss of generality, assume that Willie is observing the top link. If Willie observes a sequence from the set $\mathcal{D}$, then he knows that Alice is transmitting that particular message. Hence, we have

$$H(\vec{M}|\vec{X}_{\mathcal{W}} = \vec{x}_{\mathcal{W}}) = \begin{cases} 0 & \text{if } \vec{x}_{\mathcal{W}} \in \mathcal{D} \\ \log_2(2^n - 2^{n\delta}) & \text{if } \vec{x}_{\mathcal{W}} \notin \mathcal{D} \end{cases} \quad (60)$$

Therefore, we have

$$H(\vec{M}|\vec{X}_{\mathcal{W}}) = \left(1 - \frac{2^{n\delta}}{2^n}\right) \log_2(2^n - 2^{n\delta}). \quad (61)$$

Then, it is easy to show that

$$I(\vec{M}; \vec{X}_{\mathcal{W}}) = H_2\left(\frac{2^{n\delta}}{2^n}\right) + \frac{2^{n\delta}}{2^n} \log_2(2^{n\delta}), \quad (62)$$

$$= o(1) \quad \text{for } n \rightarrow \infty, \quad (63)$$

where $H_2(\cdot)$ denotes the binary entropy function.

Next, we show that this scheme does not satisfy the hidability criterion. For instance, for any sequence $\vec{m}_{\mathcal{D}} \in \mathcal{D}$, we will have $\Pr_{\vec{M}, Enc}(\vec{M} = \vec{m}_{\mathcal{D}}|\vec{X}_{\mathcal{W}} = \vec{m}_{\mathcal{D}}, \mathbf{T} = 1) = 1$, and thus,

$$\frac{\Pr_{\vec{M}, Enc}(\vec{M} = \vec{m}|\vec{X}_{\mathcal{W}} = \vec{m}, \mathbf{T} = 1)}{\Pr_{\vec{M}}(\vec{M} = \vec{m}|\mathbf{T} = 1)} = 2^n \quad \forall \vec{m} \in \mathcal{D}, \forall \mathcal{W} \in \mathcal{W}_{\text{all}}$$

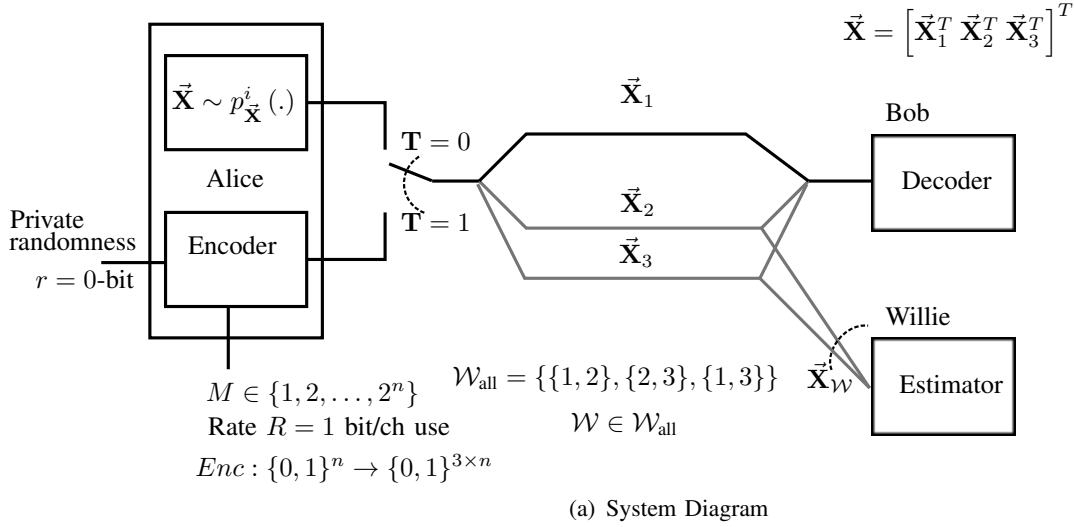
, violating the hidability. Intuitively, if Willie happens to observe a sequence from the set $\mathcal{D}$, he knows which message was transmitted, and the messages in the set $\mathcal{D}$ are not secure. One can argue that such feeble messages are asymptotically small in fraction. However, one should note that their cardinality is symptomatically very large. Moreover, looking at this from the reverse perspective clearly depicts the weakness of this non-hidable encoding scheme, as follows.

Suppose Willie is specifically interested in whether any of the message in set $\mathcal{D}$ is transmitted or not. Then, when he observes any sequence $\vec{x}_{\mathcal{W}} \notin \mathcal{D}$, he knows that no message in the set $\mathcal{D}$ was transmitted. Notice that such event occurs with a large probability of $\left(1 - \frac{2^{n\delta}}{2^n}\right)$. ■

APPENDIX B EXAMPLES ON DENIABILITY VS. HIDABILITY

In this section, we present some examples of the schemes that are hidable but not deniable and vice-versa.

First, let us consider a hidable scheme that is not deniable. For this, consider Shannon's one-time padding scheme described in the toy example in section III-B (see Fig. 2 on page 6). It is straightforward to show that this scheme is hidable. However, it induces uniform distribution on the codewords on each link. The marginal induced distribution is far away from the marginal innocent distribution, and thus, the



$\mathbf{X}(t) = [\mathbf{X}_1(t) \mathbf{X}_2(t) \mathbf{X}_3(t)]^T$	$p_{\mathbf{X}}^i(\cdot)$
0 0 0	$\frac{1}{2}$
0 0 1	0
0 1 0	0
0 1 1	0
1 0 0	0
1 0 1	0
1 1 0	0
1 1 1	$\frac{1}{2}$

(b) Scalar Innocent Distribution

$\mathbf{X}(t) = [\mathbf{X}_j(t) \mathbf{X}_k(t)]^T$	$p_{\mathbf{X}_{j,k}}^i(\cdot)$
0 0	$\frac{1}{2}$
0 1	0
1 0	0
1 1	$\frac{1}{2}$

(c) Scalar Marginal Innocent Distribution on any pair of links $\{j, k\}$

Fig. 3. Example of a parallel-link network with $C = 3$ links. Willie can observe any pair of the links.

scheme is not deniable. In fact, notice that this scheme will be deniable only if the marginal innocent distributions on both the individual links are close to uniform.

Now, consider an example of a deniable scheme that is not hidable. Consider a parallel-link network with $C = 3$ links. Willie can observe any pair of links. The scalar innocent distribution is given by Fig. 3(b). This results in the same marginal innocent distribution as given in Fig. 3(c) on each pair of links. In this case, Alice can use a simple scheme in the active state: she simply copies her cover message bit on each link. In this way, Alice can transmit one bit of covert message to Bob per network use. Since the covert message bits are assumed to be uniformly distributed, this scheme is deniable.

However, this scheme is clearly not hidable. This is because Willie can perfectly decode Alice's message from his observations, when he knows that Alice is in the active status. This makes the ratio $\frac{\Pr_{\hat{\mathbf{M}}, Enc}(\hat{\mathbf{M}}_W = \vec{\mathbf{m}} | \vec{\mathbf{x}}_W, \mathbf{T} = 1)}{\Pr_{\hat{\mathbf{M}}}(\hat{\mathbf{M}}_W = \vec{\mathbf{m}} | \mathbf{T} = 1)} = 2^n$, which is exponentially large.

APPENDIX C PROOF OF DENIABILITY AND HIDABILITY FOR MULTIPLE LINKS

Hidability: First, consider the hidability of the scheme. To achieve hidability, we want to have (see (4))

$$\frac{\Pr_{\hat{\mathbf{M}}, Enc}(\hat{\mathbf{M}} = \vec{\mathbf{m}} | \vec{\mathbf{x}}_W, \mathbf{T} = 1)}{\Pr_{\hat{\mathbf{M}}}(\hat{\mathbf{M}} = \vec{\mathbf{m}} | \mathbf{T} = 1)} \in [1 - \epsilon_h, 1 + \epsilon_h].$$

Using Bayes' rule, the above condition transforms to

$$\frac{\Pr_{\vec{\mathbf{M}}, Enc}(\vec{\mathbf{X}}_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}} | \vec{\mathbf{M}} = \vec{\mathbf{m}}, \mathbf{T} = 1)}{\Pr_{\vec{\mathbf{M}}, Enc}(\vec{\mathbf{X}}_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}} | \mathbf{T} = 1)} \in [1 - \epsilon_h, 1 + \epsilon_h]. \quad (64)$$

Let $\mathcal{W}^c = C \setminus \mathcal{W}$ denote the set of links which Willie does not observe. Let $\vec{\mathbf{x}}_{\mathcal{W}}$ be the part of the codeword $\vec{\mathbf{x}}$ that is observed by Willie. Note that $\vec{\mathbf{x}}_{\mathcal{W}} \in \mathcal{X}_{\mathcal{W}}^n$, where $\mathcal{X}_{\mathcal{W}} = \{0, 1, \dots, 2^{|\mathcal{W}|} - 1\}$. Similarly, $\vec{\mathbf{x}}_{\mathcal{W}^c}$ be the other part of the codeword $\vec{\mathbf{x}}$ that Willie cannot observe. Here, $\vec{\mathbf{x}}_{\mathcal{W}^c} \in \mathcal{X}_{\mathcal{W}^c}^n$, where $\mathcal{X}_{\mathcal{W}^c} = \{0, 1, \dots, 2^{|\mathcal{W}^c|} - 1\}$. Observe that the codeword $\vec{\mathbf{x}}$ can be decomposed as $\vec{\mathbf{x}} = [\vec{\mathbf{x}}_{\mathcal{W}}; \vec{\mathbf{x}}_{\mathcal{W}^c}]$.

Now, for a particular sequence $\vec{\mathbf{x}}_{\mathcal{W}}$ observed by Willie, we have

$$\Pr_{\vec{\mathbf{M}}, Enc}(\vec{\mathbf{X}}_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}} | \mathbf{T} = 1) \stackrel{(a)}{=} \sum_{\vec{\mathbf{x}}'_{\mathcal{W}^c}} \Pr_{\vec{\mathbf{M}}, Enc}(\vec{\mathbf{x}}_{\mathcal{W}}, \vec{\mathbf{x}}'_{\mathcal{W}^c} | \mathbf{T} = 1) \quad (65)$$

$$\stackrel{(b)}{=} \sum_{\vec{\mathbf{m}} \in \vec{\mathbf{M}}} \sum_{\vec{\mathbf{x}}': \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}}} \Pr_{\vec{\mathbf{M}}, Enc}(\vec{\mathbf{x}}', \vec{\mathbf{m}} | \mathbf{T} = 1) \quad (66)$$

$$= \sum_{\vec{\mathbf{m}} \in \vec{\mathbf{M}}} \sum_{\vec{\mathbf{x}}': \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}}} \Pr_{\vec{\mathbf{M}}, Enc}(\vec{\mathbf{x}}' | \vec{\mathbf{m}}, \mathbf{T} = 1) \Pr_{\vec{\mathbf{M}}, Enc}(\vec{\mathbf{m}} | \mathbf{T} = 1) \quad (67)$$

$$= \sum_{\vec{\mathbf{m}} \in \vec{\mathbf{M}}} \sum_{\vec{\mathbf{x}}': \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}}} p_{\vec{\mathbf{X}} | \vec{\mathbf{M}}}^{\text{stoch}}(\vec{\mathbf{x}}' | \vec{\mathbf{m}}, \mathbf{T} = 1) \Pr_{\vec{\mathbf{M}}, Enc}(\vec{\mathbf{m}} | \mathbf{T} = 1) \quad (68)$$

$$\stackrel{(c)}{=} \sum_{\vec{\mathbf{m}} \in \vec{\mathbf{M}}} \sum_{\substack{\vec{\mathbf{x}}': \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}} \\ \vec{\mathbf{x}}' \in \mathcal{B}(\vec{\mathbf{m}})}} \frac{p_{\vec{\mathbf{X}}}^i(\vec{\mathbf{x}}')}{\Pr_{\mathcal{B}}(\mathcal{B}(\vec{\mathbf{m}}))} \Pr_{\vec{\mathbf{M}}, Enc}(\vec{\mathbf{m}} | \mathbf{T} = 1), \quad (69)$$

where (a), (b) follow from the rule of total probability, and (c) follows from the definition of stochastic encoder given in (39). Using the result in lemma 1, we can bound the probability $\Pr_{\mathcal{B}}(\mathcal{B}(\vec{\mathbf{m}}))$, and thus, we have

$$\begin{aligned} \sum_{\vec{\mathbf{m}} \in \vec{\mathbf{M}}} \sum_{\substack{\vec{\mathbf{x}}': \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}} \\ \vec{\mathbf{x}}' \in \mathcal{B}(\vec{\mathbf{m}})}} \frac{p_{\vec{\mathbf{X}}}^i(\vec{\mathbf{x}}')}{(1 + \epsilon_1) \Pr_{\vec{\mathbf{M}}, Enc}(\vec{\mathbf{m}} | \mathbf{T} = 1)} \Pr_{\vec{\mathbf{M}}, Enc}(\vec{\mathbf{m}} | \mathbf{T} = 1) &\leq \Pr_{\vec{\mathbf{M}}, Enc}(\vec{\mathbf{X}}_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}} | \mathbf{T} = 1) \\ &\leq \sum_{\vec{\mathbf{m}} \in \vec{\mathbf{M}}} \sum_{\substack{\vec{\mathbf{x}}': \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}} \\ \vec{\mathbf{x}}' \in \mathcal{B}(\vec{\mathbf{m}})}} \frac{p_{\vec{\mathbf{X}}}^i(\vec{\mathbf{x}}')}{(1 - \epsilon_1) \Pr_{\vec{\mathbf{M}}, Enc}(\vec{\mathbf{m}} | \mathbf{T} = 1)} \Pr_{\vec{\mathbf{M}}, Enc}(\vec{\mathbf{m}} | \mathbf{T} = 1) \end{aligned} \quad (70)$$

$$(71)$$

for some $\epsilon_1 > 0$. Simplifying, we get

$$\frac{1}{(1 + \epsilon_1)} \sum_{\substack{\vec{\mathbf{x}}': \vec{\mathbf{x}}' \in \mathcal{T}_{\epsilon}^{(n)} \\ \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}}}} p_{\vec{\mathbf{X}}}^i(\vec{\mathbf{x}}') \leq \Pr_{\vec{\mathbf{M}}, Enc}(\vec{\mathbf{X}}_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}} | \mathbf{T} = 1) \leq \frac{1}{(1 - \epsilon_1)} \sum_{\substack{\vec{\mathbf{x}}': \vec{\mathbf{x}}' \in \mathcal{T}_{\epsilon}^{(n)} \\ \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}}}} p_{\vec{\mathbf{X}}}^i(\vec{\mathbf{x}}'). \quad (72)$$

We slightly abuse the notation to write the above inequalities as follows:

$$\Pr_{\vec{\mathbf{M}}, Enc}(\vec{\mathbf{X}}_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}} | \mathbf{T} = 1) \in \frac{1}{[1 + \epsilon_1, 1 - \epsilon_1]} \sum_{\substack{\vec{\mathbf{x}}': \vec{\mathbf{x}}' \in \mathcal{T}_{\epsilon}^{(n)} \\ \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}}}} p_{\vec{\mathbf{X}}}^i(\vec{\mathbf{x}}'). \quad (73)$$

Note that, as proved in Lemma 3 in appendix D, for any strongly typical $\vec{x}$ (w.r.t. $p_{\vec{x}}^i(\cdot)$), the sequence $\vec{x}_{\mathcal{W}}$ observed by Willie is also strongly typical (w.r.t. $p_{\vec{x}_{\mathcal{W}}}^i(\cdot)$). Furthermore, the sequence not observed by Willie, $\vec{x}_{\mathcal{W}^c}$, is conditionally strongly typical given $\vec{x}_{\mathcal{W}}$ (Lemma 3). Thus, the set $\left\{ \vec{x}' : \vec{x}' \in \mathcal{T}_{\epsilon}^{(n)}, \vec{x}'_{\mathcal{W}} = \vec{x}_{\mathcal{W}} \right\}$ is the same as the conditionally strongly typical set $\mathcal{T}_{\epsilon}^{(n)}(\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}})$ (w.r.t. $p_{\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}}}^i(\cdot|\cdot)$) defined as follows:

$$\mathcal{T}_{\epsilon}^{(n)}(\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}}) = \left\{ \vec{x}_{\mathcal{W}^c} : \begin{cases} \left| \frac{1}{n} N(a, b|\vec{x}_{\mathcal{W}}, \vec{x}_{\mathcal{W}^c}) - p_{\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}}}^i(b|a) N(a|\vec{x}_{\mathcal{W}}) \right| \leq \frac{\epsilon(1+|\mathcal{X}_{\mathcal{W}^c}|)}{|\mathcal{X}_{\mathcal{W}}||\mathcal{X}_{\mathcal{W}^c}|} \\ \forall (a, b) \in \mathcal{X}_{\mathcal{W}} \times \mathcal{X}_{\mathcal{W}^c}, & \text{if } p_{\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}}}^i(b|a) > 0 \\ N(a, b|\vec{x}_{\mathcal{W}}, \vec{x}_{\mathcal{W}^c}) = 0, & \text{if } p_{\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}}}^i(b|a) = 0 \end{cases} \right\}. \quad (74)$$

Therefore, the equations (73) can be written as

$$\begin{aligned} & \Pr_{\vec{M}, Enc}(\vec{X}_{\mathcal{W}} = \vec{x}_{\mathcal{W}} | \mathbf{T} = 1) \\ & \in \frac{1}{[1 + \epsilon_1, 1 - \epsilon_1]} \sum_{\vec{x}': \vec{x}'_{\mathcal{W}^c} \in \mathcal{T}_{\epsilon}^{(n)}(\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}})} p_{\vec{x}}^i(\vec{x}') \end{aligned} \quad (75)$$

$$\stackrel{(a)}{\in} \frac{1}{[1 + \epsilon_1, 1 - \epsilon_1]} \sum_{Q_{\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}}} \in \mathcal{T}_{\epsilon}^{(n)}(\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}})} \sum_{\vec{x}': \vec{x}'_{\mathcal{W}^c} \in Q_{\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}}}} p_{\vec{x}}^i(\vec{x}') \quad (76)$$

$$\stackrel{(b)}{\in} \frac{1}{[1 + \epsilon_1, 1 - \epsilon_1]} \sum_{Q_{\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}}} \in \mathcal{T}_{\epsilon}^{(n)}(\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}})} |Q_{\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}}}| p_{\vec{x}}^i(\vec{x}'_{Q_{\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}}}}), \quad (77)$$

where $\vec{x}'_{Q_{\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}}}}$ denotes a vector $\vec{x}' : \vec{x}'_{\mathcal{W}^c} \in Q_{\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}}}$. Further, (a) follows by partitioning the conditionally strongly typical set into the conditional type classes $Q_{\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}}}$ corresponding to conditional types $Q_{\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}}}(b|a) = \frac{N(a, b|\vec{x}_{\mathcal{W}}, \vec{x}_{\mathcal{W}^c})}{N(a|\vec{x}_{\mathcal{W}})}$ for $(a, b) \in \mathcal{X}_{\mathcal{W}} \times \mathcal{X}_{\mathcal{W}^c}$. For step (b), note that for a given strongly typical $\vec{x}_{\mathcal{W}}$, the probability of all the codewords $\vec{x}'$ such that $\vec{x}'_{\mathcal{W}^c}$ lies in a particular conditional type class is the same.

Next, consider the conditional probability of Willie observing a particular sequence $\vec{x}_{\mathcal{W}}$ as follows:

$$\begin{aligned} & \Pr_{\vec{M}, Enc}(\vec{X}_{\mathcal{W}} = \vec{x}_{\mathcal{W}} | \vec{M} = \vec{m}, \mathbf{T} = 1) \\ & \stackrel{(a)}{=} \sum_{\vec{x}'_{\mathcal{W}^c}} \Pr_{\vec{M}, Enc}(\vec{x}_{\mathcal{W}}, \vec{x}'_{\mathcal{W}^c} | \vec{M} = \vec{m}, \mathbf{T} = 1) \end{aligned} \quad (78)$$

$$= \sum_{\vec{x}': \vec{x}'_{\mathcal{W}} = \vec{x}_{\mathcal{W}}} \Pr_{\vec{M}, Enc}(\vec{x}' | \vec{M} = \vec{m}, \mathbf{T} = 1) \quad (79)$$

$$= \sum_{\vec{x}': \vec{x}'_{\mathcal{W}} = \vec{x}_{\mathcal{W}}} p_{\vec{x}|\vec{M}}^{\text{stoch}}(\vec{x}' | \vec{m}, \mathbf{T} = 1) \quad (80)$$

$$\stackrel{(b)}{=} \sum_{\substack{\vec{x}': \vec{x}'_{\mathcal{W}} = \vec{x}_{\mathcal{W}} \\ \vec{x}' \in \mathcal{B}(\vec{m})}} \frac{p_{\vec{x}}^i(\vec{x}')}{\Pr_{\mathcal{B}}(\mathcal{B}(\vec{m}))}, \quad (81)$$

where (a) follows from the rule of total probability and (b) follows from the definition of the stochastic

encoder given by (39). Again, using lemma 1 to bound $\Pr_{\mathcal{B}}(\mathcal{B}(\vec{\mathbf{m}}))$, we get

$$\begin{aligned} \sum_{\substack{\vec{\mathbf{x}}': \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}} \\ \vec{\mathbf{x}}' \in \mathcal{B}(\vec{\mathbf{m}})}} \frac{p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}}')}{(1 + \epsilon_1) \Pr_{\vec{\mathbf{M}}, Enc}(\vec{\mathbf{m}}|\mathbf{T} = 1)} &\leq \Pr_{\vec{\mathbf{M}}, Enc}(\vec{\mathbf{X}}_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}}|\vec{\mathbf{M}} = \vec{\mathbf{m}}, \mathbf{T} = 1) \\ &\leq \sum_{\substack{\vec{\mathbf{x}}': \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}} \\ \vec{\mathbf{x}}' \in \mathcal{B}(\vec{\mathbf{m}})}} \frac{p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}}')}{(1 - \epsilon_1) \Pr_{\vec{\mathbf{M}}, Enc}(\vec{\mathbf{m}}|\mathbf{T} = 1)} \end{aligned} \quad (82)$$

$$(83)$$

Since all the codewords $\vec{\mathbf{x}}'$ are strongly typical, we have

$$\begin{aligned} \frac{1}{\Pr_{\vec{\mathbf{M}}}(\vec{\mathbf{m}}|\mathbf{T} = 1)(1 + \epsilon_1)} \sum_{\substack{\vec{\mathbf{x}}': \vec{\mathbf{x}}' \in \mathcal{T}_{\epsilon}^{(n)} \cap \mathcal{B}(\vec{\mathbf{m}}) \\ \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}}}} p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}}') &\leq \Pr_{\vec{\mathbf{M}}, Enc}(\vec{\mathbf{X}}_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}}|\vec{\mathbf{M}} = \vec{\mathbf{m}}, \mathbf{T} = 1) \\ &\leq \frac{1}{\Pr_{\vec{\mathbf{M}}}(\vec{\mathbf{m}}|\mathbf{T} = 1)(1 - \epsilon_1)} \sum_{\substack{\vec{\mathbf{x}}': \vec{\mathbf{x}}' \in \mathcal{T}_{\epsilon}^{(n)} \cap \mathcal{B}(\vec{\mathbf{m}}) \\ \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}}}} p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}}'). \end{aligned} \quad (84)$$

Again, abusing the notation, we write the above inequalities as follows

$$\Pr_{\vec{\mathbf{M}}, Enc}(\vec{\mathbf{X}}_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}}|\vec{\mathbf{M}} = \vec{\mathbf{m}}, \mathbf{T} = 1) \in \frac{1}{\Pr_{\vec{\mathbf{M}}}(\vec{\mathbf{m}}|\mathbf{T} = 1) [1 + \epsilon_1, 1 - \epsilon_1]} \sum_{\substack{\vec{\mathbf{x}}': \vec{\mathbf{x}}' \in \mathcal{T}_{\epsilon}^{(n)} \cap \mathcal{B}(\vec{\mathbf{m}}) \\ \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}}}} p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}}') \quad (85)$$

Using the fact that the set $\{\vec{\mathbf{x}}' : \vec{\mathbf{x}}' \in \mathcal{T}_{\epsilon}^{(n)}, \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}}\}$ is the same as the conditionally strongly typical set $\mathcal{T}_{\epsilon}^{(n)}(\vec{\mathbf{X}}_{\mathcal{W}^c}|\vec{\mathbf{x}}_{\mathcal{W}})$ (w.r.t. $p_{\mathbf{X}_{\mathcal{W}^c}|\mathbf{X}_{\mathcal{W}}}^i(\cdot|\cdot)$), it is easy to show that

$$\begin{aligned} \Pr_{\vec{\mathbf{M}}, Enc}(\vec{\mathbf{X}}_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}}|\vec{\mathbf{M}} = \vec{\mathbf{m}}, \mathbf{T} = 1) &\in \frac{1}{\Pr_{\vec{\mathbf{M}}}(\vec{\mathbf{m}}|\mathbf{T} = 1) [1 + \epsilon_1, 1 - \epsilon_1]} \sum_{Q_{\vec{\mathbf{x}}_{\mathcal{W}^c}|\vec{\mathbf{x}}_{\mathcal{W}}} \in \mathcal{T}_{\epsilon}^{(n)}(\vec{\mathbf{x}}_{\mathcal{W}^c}|\vec{\mathbf{x}}_{\mathcal{W}})} |Q_{\vec{\mathbf{x}}_{\mathcal{W}^c}|\vec{\mathbf{x}}_{\mathcal{W}}} \cap \mathcal{B}(\vec{\mathbf{m}})| p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}}'_{Q_{\vec{\mathbf{x}}_{\mathcal{W}^c}|\vec{\mathbf{x}}_{\mathcal{W}}}}), \end{aligned} \quad (86)$$

where $\vec{\mathbf{x}}'_{Q_{\vec{\mathbf{x}}_{\mathcal{W}^c}|\vec{\mathbf{x}}_{\mathcal{W}}}}$ denotes a vector $\vec{\mathbf{x}}' : \vec{\mathbf{x}}'_{\mathcal{W}^c} \in Q_{\vec{\mathbf{x}}_{\mathcal{W}^c}|\vec{\mathbf{x}}_{\mathcal{W}}}$.

Now, by using lemma 4 in appendix D, if $|\vec{\mathbf{M}}| \leq 2^{n(H(p_{\mathbf{X}_{\mathcal{W}^c}|\mathbf{X}_{\mathcal{W}}}^i(\cdot|\cdot)) - \tilde{r}_d^1)}$ such that $\tilde{r}_d^1 \geq 2\epsilon \log_2 \frac{2^C}{\epsilon} + \frac{2^C \log_2(n+1)}{n}$, then with high probability

$$\begin{aligned} \Pr_{\vec{\mathbf{M}}, Enc}(\vec{\mathbf{X}}_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}}|\vec{\mathbf{M}} = \vec{\mathbf{m}}, \mathbf{T} = 1) &\in \frac{1}{\Pr_{\vec{\mathbf{M}}}(\vec{\mathbf{m}}|\mathbf{T} = 1) [1 + \epsilon_1, 1 - \epsilon_1]} \sum_{Q_{\vec{\mathbf{x}}_{\mathcal{W}^c}|\vec{\mathbf{x}}_{\mathcal{W}}} \in \mathcal{T}_{\epsilon}^{(n)}(\vec{\mathbf{x}}_{\mathcal{W}^c}|\vec{\mathbf{x}}_{\mathcal{W}})} \frac{|Q_{\vec{\mathbf{x}}_{\mathcal{W}^c}|\vec{\mathbf{x}}_{\mathcal{W}}|}{|\vec{\mathbf{M}}|} p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}}'_{Q_{\vec{\mathbf{x}}_{\mathcal{W}^c}|\vec{\mathbf{x}}_{\mathcal{W}}}}) \end{aligned} \quad (87)$$

$$\stackrel{(a)}{\in} \frac{1}{[1 + \epsilon_1, 1 - \epsilon_1]} \sum_{Q_{\vec{\mathbf{x}}_{\mathcal{W}^c}|\vec{\mathbf{x}}_{\mathcal{W}}} \in \mathcal{T}_{\epsilon}^{(n)}(\vec{\mathbf{x}}_{\mathcal{W}^c}|\vec{\mathbf{x}}_{\mathcal{W}})} |Q_{\vec{\mathbf{x}}_{\mathcal{W}^c}|\vec{\mathbf{x}}_{\mathcal{W}}| p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}}'_{Q_{\vec{\mathbf{x}}_{\mathcal{W}^c}|\vec{\mathbf{x}}_{\mathcal{W}}}}), \quad (88)$$

where (a) follows from $\Pr_{\vec{\mathbf{M}}}(\vec{\mathbf{m}}|\mathbf{T} = 1) = \frac{1}{|\vec{\mathbf{M}}|}$.

Finally, from (77) and (88), we have

$$\frac{1 - \epsilon_1}{1 + \epsilon_1} \leq \frac{\Pr_{\vec{\mathbf{M}}, Enc}(\vec{\mathbf{X}}_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}} | \vec{\mathbf{M}} = \vec{\mathbf{m}}, \mathbf{T} = 1)}{\Pr_{\vec{\mathbf{M}}, Enc}(\vec{\mathbf{X}}_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}} | \mathbf{T} = 1)} \leq \frac{1 + \epsilon_1}{1 - \epsilon_1} \quad (89)$$

with high probability if $R_s \leq H(p_{\mathbf{X}_{\mathcal{W}^c}|\mathbf{X}_{\mathcal{W}}}^i(\cdot|\cdot)) - \tilde{r}_d^1$, where $\tilde{r}_d^1 > 2\epsilon \log_2 \frac{2^C}{\epsilon} + \frac{2^C \log_2(n+1)}{n}$, and hence the proof.

Deniability: Next, consider the deniability of the proposed scheme. As proved in theorem 1, random binning based procedure will be deniable provided $R_s \leq H(p_{\mathbf{X}}^i(\cdot)) - \tilde{r}_d$, where $\tilde{r}_d > \epsilon \log_2 \frac{2^C}{\epsilon} + \frac{2^C \log_2(n+1)}{n}$. Note that, since conditioning reduces entropy, $H(p_{\mathbf{X}_{\mathcal{W}^c}|\mathbf{X}_{\mathcal{W}}}^i(\cdot|\cdot)) - \tilde{r}_d^1 \leq H(p_{\mathbf{X}}^i(\cdot)) - \tilde{r}_d$. Therefore, the scheme is deniable.

Reliability: We modify the encoding procedure as mentioned in appendix E to achieve reliability. The proof goes on the exact similar lines as in appendix E.

APPENDIX D USEFUL LEMMAS

First, we show that, under random binning based encoding schemes, the probability of any message bin is bounded.

Lemma 1. *In the random binning based stochastic encoding schemes (Procedures 1 and 2), the probability of a message bin $\Pr_{\mathcal{B}}(\mathcal{B}(\vec{\mathbf{m}})) = \sum_{\vec{\mathbf{x}}: \vec{\mathbf{x}} \in \mathcal{C} \cap \mathcal{B}(\vec{\mathbf{m}})} p_{\vec{\mathbf{x}}}^a(\vec{\mathbf{x}})$ is bounded as*

$$(1 - \epsilon) \left(1 - 2^{C+1} \exp(-2\epsilon^2 n)\right) \frac{1}{|\vec{\mathbf{M}}|} \leq \Pr_{\mathcal{B}}(\mathcal{B}(\vec{\mathbf{m}})) \leq (1 + \epsilon) \frac{1}{|\vec{\mathbf{M}}|}. \quad (90)$$

with probability at least

$$1 - 2 \exp\left(\frac{-\epsilon^2 2^n \left[H(Q) - \frac{2^C \log_2(n+1)}{n}\right]}{3 |\vec{\mathbf{M}}|}\right),$$

where $|\vec{\mathbf{M}}|$ denotes the total number of covert messages.

Proof: Beginning with the definition of the probability of a message bin, we can write

$$\Pr_{\mathcal{B}}(\mathcal{B}(\vec{\mathbf{m}})) = \sum_{\vec{\mathbf{x}}: \vec{\mathbf{x}} \in \mathcal{C} \cap \mathcal{B}(\vec{\mathbf{m}})} p_{\vec{\mathbf{x}}}^a(\vec{\mathbf{x}}), \quad (91)$$

$$\stackrel{(a)}{=} \sum_{\vec{\mathbf{x}} \in \mathcal{T}_{\epsilon}^{(n)} \cap \mathcal{B}(\vec{\mathbf{m}})} p_{\vec{\mathbf{x}}}^a(\vec{\mathbf{x}}), \quad (92)$$

$$\stackrel{(b)}{=} \sum_{Q \in \mathcal{T}_{\epsilon}^{(n)}} \sum_{\vec{\mathbf{x}} \in Q \cap \mathcal{B}(\vec{\mathbf{m}})} p_{\vec{\mathbf{x}}}^a(\vec{\mathbf{x}}) \quad (93)$$

$$\stackrel{(c)}{=} \sum_{Q \in \mathcal{T}_{\epsilon}^{(n)}} |Q \cap \mathcal{B}(\vec{\mathbf{m}})| p_{\vec{\mathbf{x}}}^a(\vec{\mathbf{x}} : \vec{\mathbf{x}} \in Q), \quad (94)$$

where (a) follows since in the random binning based schemes, the codebook is the ϵ -strongly typical set, i.e., $\mathcal{C} = \mathcal{T}_{\epsilon}^{(n)}$, (b) follows because the strongly typical set can be considered as a collection of several type classes, and (c) follows from the fact that the probability of each codeword in any particular type class is the same. Notice that, here, $|Q \cap \mathcal{B}(\vec{\mathbf{m}})|$ denote the number of codewords of Q that are associated with the bin $\mathcal{B}(\vec{\mathbf{m}})$.

Taking expectation of both sides with respect to the binning process, we get

$$\mathbb{E}_{\mathcal{B}} [\Pr_{\mathcal{B}} (\mathcal{B}(\vec{\mathbf{m}}))] = \sum_{Q \in \mathcal{T}_{\epsilon}^{(n)}} \mathbb{E}_{\mathcal{B}} [|Q \cap \mathcal{B}(\vec{\mathbf{m}})|] p_{\vec{\mathbf{x}}}^a (\vec{\mathbf{x}} : \vec{\mathbf{x}} \in Q) \quad (95)$$

$$\stackrel{(d)}{=} \sum_{Q \in \mathcal{T}_{\epsilon}^{(n)}} \frac{|Q|}{|\vec{\mathbf{M}}|} p_{\vec{\mathbf{x}}}^a (\vec{\mathbf{x}} : \vec{\mathbf{x}} \in Q) \quad (96)$$

$$= \frac{1}{|\vec{\mathbf{M}}|} \sum_{Q \in \mathcal{T}_{\epsilon}^{(n)}} |Q| p_{\vec{\mathbf{x}}}^a (\vec{\mathbf{x}} : \vec{\mathbf{x}} \in Q_j) \quad (97)$$

$$\stackrel{(e)}{=} \frac{1}{|\vec{\mathbf{M}}|} \Pr_{\vec{\mathbf{x}}} (\mathcal{T}_{\epsilon}^{(n)} (p_{\vec{\mathbf{x}}}^a (\cdot))) \quad (98)$$

$$(99)$$

where, to get (d), recall that in a random binning experiment with m balls and k bins, the expected number of balls in a bin is $\frac{m}{k}$. For (e), observe that $|Q| p_{\vec{\mathbf{x}}}^a (\vec{\mathbf{x}} : \vec{\mathbf{x}} \in Q) = \Pr (Q)$. Finally, using the following lower bound on the probability of ϵ -strongly typical set $\Pr (\mathcal{T}_{\epsilon}^{(n)}) \geq 1 - 2 |\mathcal{X}| \exp (-2\epsilon^2 n)$ (see [12]), we have

$$\mathbb{E}_{\mathcal{B}} [\Pr_{\mathcal{B}} (\mathcal{B}(\vec{\mathbf{m}}))] \in (1 - 2^{C+1} \exp (-2\epsilon^2 n), 1) \frac{1}{|\vec{\mathbf{M}}|}, \quad (100)$$

where we substitute $|\mathcal{X}| = 2^C$.

Now, consider

$$|\Pr_{\mathcal{B}} (\mathcal{B}(\vec{\mathbf{m}})) - \mathbb{E}_{\mathcal{B}} [\Pr_{\mathcal{B}} (\mathcal{B}(\vec{\mathbf{m}}))]| \stackrel{(g)}{=} \left| \sum_{Q \in \mathcal{T}_{\epsilon}^{(n)}} |Q \cap \mathcal{B}(\vec{\mathbf{m}})| p_{\vec{\mathbf{x}}}^a (\vec{\mathbf{x}} : \vec{\mathbf{x}} \in Q) \right. \quad (101)$$

$$\left. - \sum_{Q \in \mathcal{T}_{\epsilon}^{(n)}} \mathbb{E}_{\mathcal{B}} [|Q \cap \mathcal{B}(\vec{\mathbf{m}})|] p_{\vec{\mathbf{x}}}^a (\vec{\mathbf{x}} : \vec{\mathbf{x}} \in Q) \right|, \quad (102)$$

$$\leq \sum_{Q \in \mathcal{T}_{\epsilon}^{(n)}} \left| |Q \cap \mathcal{B}(\vec{\mathbf{m}})| - \mathbb{E}_{\mathcal{B}} [|Q \cap \mathcal{B}(\vec{\mathbf{m}})|] \right| p_{\vec{\mathbf{x}}}^a (\vec{\mathbf{x}} : \vec{\mathbf{x}} \in Q), \quad (103)$$

$$\stackrel{(h)}{\leq} \sum_{Q \in \mathcal{T}_{\epsilon}^{(n)}} \epsilon \mathbb{E}_{\mathcal{B}} [|Q \cap \mathcal{B}(\vec{\mathbf{m}})|] p_{\vec{\mathbf{x}}}^a (\vec{\mathbf{x}} : \vec{\mathbf{x}} \in Q), \quad \text{w.h.p.} \quad (104)$$

$$\stackrel{(i)}{=} \epsilon \mathbb{E}_{\mathcal{B}} [\Pr_{\mathcal{B}} (\mathcal{B}(\vec{\mathbf{m}}))], \quad (105)$$

$$\therefore \Pr_{\mathcal{B}} (\mathcal{B}(\vec{\mathbf{m}})) \in [(1 - \epsilon), (1 + \epsilon)] \mathbb{E}_{\mathcal{B}} [\Pr_{\mathcal{B}} (\mathcal{B}(\vec{\mathbf{m}}))], \quad \text{w.h.p.} \quad (106)$$

where (g) follows from (94), (h) follows from Lemma 2 proven below, and (i) follows again from (94). To be precise, note that we used (107) to obtain inequality (h), and thus, equation (106) holds with probability at least $1 - 2 \exp \left(\frac{-\epsilon^2 2^n [H(Q) - 2^C \log_2(n+1)]}{3|\vec{\mathbf{M}}|} \right)$, with $|\vec{\mathbf{M}}|$ being the total number of covert messages. Finally, the result (90) follows from (100) and (106). ■

In the following lemma, we show that, under the random binning based encoding scheme, the number of codewords of any type class per message bin is concentrated around its mean value.

Lemma 2. *In the random binning based stochastic encoding schemes (Procedures 1 and 2), for a type class Q , we have*

$$\Pr \left(\left| |Q \cap \mathcal{B}(\vec{\mathbf{m}})| - \mathbb{E}_{\mathcal{B}} [|Q \cap \mathcal{B}(\vec{\mathbf{m}})|] \right| \geq \epsilon \mathbb{E}_{\mathcal{B}} [|Q \cap \mathcal{B}(\vec{\mathbf{m}})|] \right) \leq 2 \exp \left(\frac{-\epsilon^2 2^n \left[H(Q) - \frac{2^C \log_2(n+1)}{n} \right]}{3 |\vec{\mathbf{M}}|} \right). \quad (107)$$

Proof: Recall that $\mathbb{E}_{\mathcal{B}} [|Q \cap \mathcal{B}(\vec{\mathbf{m}})|]$ denote the average number of codewords of type class Q that are associated with the bin $\mathcal{B}(\vec{\mathbf{m}})$. For any particular type class Q , we have $|Q|$ codewords which will be associated uniformly at random with $|\vec{\mathbf{M}}|$ message bins. Therefore, the expected number of codewords of a particular type class which fall into a given message bin is $\mathbb{E}_{\mathcal{B}} [|Q \cap \mathcal{B}(\vec{\mathbf{m}})|] = \frac{|Q|}{|\vec{\mathbf{M}}|}$. Moreover, note that for any type (empirical distribution) Q defined over an alphabet $\mathcal{X}$, the type class Q satisfies (see [13])

$$\frac{1}{(n+1)^{|\mathcal{X}|}} 2^{nH(Q)} \leq |Q| \leq 2^{nH(Q)} \quad (108)$$

Hence, using $|\mathcal{X}| = 2^C$, we get

$$\frac{1}{(n+1)^{2^C}} \frac{2^{nH(Q)}}{|\vec{\mathbf{M}}|} \leq \mathbb{E}_{\mathcal{B}} [|Q \cap \mathcal{B}(\vec{\mathbf{m}})|] \leq \frac{2^{nH(Q)}}{|\vec{\mathbf{M}}|}. \quad (109)$$

Then, by using Chernoff bound, we can write

$$\begin{aligned} \Pr \left(\left| |Q \cap \mathcal{B}(\vec{\mathbf{m}})| - \mathbb{E}_{\mathcal{B}} [|Q \cap \mathcal{B}(\vec{\mathbf{m}})|] \right| \geq \epsilon \mathbb{E}_{\mathcal{B}} [|Q \cap \mathcal{B}(\vec{\mathbf{m}})|] \right) \\ \leq 2 \exp \left(\frac{-\epsilon^2 \mathbb{E}_{\mathcal{B}} [|Q \cap \mathcal{B}(\vec{\mathbf{m}})|]}{3} \right), \end{aligned} \quad (110)$$

$$\stackrel{(a)}{\leq} 2 \exp \left(\frac{-\epsilon^2 2^{nH(Q)}}{3(n+1)^{2^C} |\vec{\mathbf{M}}|} \right), \quad (111)$$

where (a) follows from the lower bound in (109). The result follows after simplifying (111). $\blacksquare$

We consider how the strong typicality of the transmitted codeword imposes the typicality on the sequence that is observed by and not observed by Willie.

Lemma 3. *Suppose the transmitted codeword $\vec{\mathbf{x}}$ is strongly typical w.r.t. the distribution $p_{\mathbf{X}}^i(\mathbf{x})$. Then, the sequence observed by Willie is strongly typical w.r.t. the marginal distribution $p_{\mathbf{X}_{\mathcal{W}}}^i(\mathbf{x}_{\mathcal{W}})$. Further, this in turn implies that the sequence that Willie cannot observe is conditionally strongly typical given $\vec{\mathbf{x}}_{\mathcal{W}}$ w.r.t. the conditional distribution $p_{\mathbf{X}_{\mathcal{W}^c}|\mathbf{X}_{\mathcal{W}}}^i(\mathbf{x}_{\mathcal{W}^c}|\mathbf{x}_{\mathcal{W}})$.*

Proof: Note that the transmitted codeword can be decomposed into the part observed by Willie and the part not observed by Willie as $\vec{\mathbf{x}} = [\vec{\mathbf{x}}_{\mathcal{W}}; \vec{\mathbf{x}}_{\mathcal{W}^c}]$, where $\vec{\mathbf{x}} \in \mathcal{X}^n = \{0, 1, \dots, 2^C - 1\}^n$, $\vec{\mathbf{x}}_{\mathcal{W}} \in \mathcal{X}_{\mathcal{W}}^n = \{0, 1, \dots, 2^{|\mathcal{W}|} - 1\}^n$, and $\vec{\mathbf{x}}_{\mathcal{W}^c} \in \mathcal{X}_{\mathcal{W}^c}^n = \{0, 1, \dots, 2^{|\mathcal{W}^c|} - 1\}^n$. Further, each symbol $c \in \mathcal{X}$ of the codeword is associated with a pair of symbols $(a, b) \in \mathcal{X}_{\mathcal{W}} \times \mathcal{X}_{\mathcal{W}^c}$.

Now, for a strongly typical codeword $\vec{\mathbf{x}} \in \mathcal{T}_{\epsilon}^{(n)}$, we have

$$\begin{aligned} \left| \frac{1}{n} N(c|\vec{\mathbf{x}}) - p_{\mathbf{X}}^i(c) \right| &\leq \frac{\epsilon}{|\mathcal{X}|} \quad \forall c \in \mathcal{X} \\ \therefore \left| \frac{1}{n} N(a, b|\vec{\mathbf{x}}_{\mathcal{W}}, \vec{\mathbf{x}}_{\mathcal{W}^c}) - p_{\mathbf{X}_{\mathcal{W}}, \mathbf{X}_{\mathcal{W}^c}}^i(a, b) \right| &\leq \frac{\epsilon}{|\mathcal{X}_{\mathcal{W}}| |\mathcal{X}_{\mathcal{W}^c}|} \quad \forall (a, b) \in \mathcal{X}_{\mathcal{W}} \times \mathcal{X}_{\mathcal{W}^c}, \end{aligned} \quad (112)$$

where distribution $p_{\mathbf{X}}^i(\cdot)$ can be represented as the joint distribution $p_{\mathbf{X}_{\mathcal{W}}, \mathbf{X}_{\mathcal{W}^c}}^i(\cdot, \cdot)$. Then, for a sequence $\vec{\mathbf{x}}_{\mathcal{W}}$ observed by Willie, we have

$$\begin{aligned} \left| \frac{1}{n} N(a|\vec{\mathbf{x}}_{\mathcal{W}}) - p_{\mathbf{X}_{\mathcal{W}}}^i(a) \right| &= \left| \sum_{b \in \mathcal{X}_{\mathcal{W}^c}} \frac{1}{n} N(a, b|\vec{\mathbf{x}}_{\mathcal{W}}, \vec{\mathbf{x}}_{\mathcal{W}^c}) - \sum_{b \in \mathcal{X}_{\mathcal{W}^c}} p_{\mathbf{X}}^i(a, b) \right| \quad \forall a \in \mathcal{X}_{\mathcal{W}} \\ &\leq \sum_{b \in \mathcal{X}_{\mathcal{W}^c}} \left| \frac{1}{n} N(a, b|\vec{\mathbf{x}}_{\mathcal{W}}, \vec{\mathbf{x}}_{\mathcal{W}^c}) - p_{\mathbf{X}_{\mathcal{W}}, \mathbf{X}_{\mathcal{W}^c}}^i(a, b) \right| \quad \forall a \in \mathcal{X}_{\mathcal{W}} \\ &\stackrel{(p)}{\leq} \frac{\epsilon}{|\mathcal{X}_{\mathcal{W}}|} \quad \forall a \in \mathcal{X}_{\mathcal{W}}, \end{aligned} \quad (113)$$

where (p) follows from the strong typicality of $\vec{\mathbf{x}}$ (see (112)). Hence, when $\vec{\mathbf{x}}$ is strongly typical w.r.t. $p_{\mathbf{X}}^i(\cdot)$, $\vec{\mathbf{x}}_{\mathcal{W}}$ is strongly typical w.r.t. $p_{\mathbf{X}_{\mathcal{W}}}^i(\cdot)$.

Now, in order to prove the conditional strong typicality of $\vec{\mathbf{x}}_{\mathcal{W}^c}$ given $\vec{\mathbf{x}}_{\mathcal{W}}$, observe that

$$\begin{aligned} \frac{1}{n} \left| N(a, b|\vec{\mathbf{x}}_{\mathcal{W}}, \vec{\mathbf{x}}_{\mathcal{W}^c}) - p_{\mathbf{X}_{\mathcal{W}^c}|\mathbf{X}_{\mathcal{W}}}^i(b|a) N(a|\vec{\mathbf{x}}_{\mathcal{W}}) \right| \\ \stackrel{(q)}{\leq} \frac{\epsilon}{|\mathcal{X}_{\mathcal{W}}|} \left(\frac{1}{|\mathcal{X}_{\mathcal{W}^c}|} + p_{\mathbf{X}_{\mathcal{W}^c}|\mathbf{X}_{\mathcal{W}}}^i(b|a) \right) \quad \forall (a, b) \in \mathcal{X}_{\mathcal{W}} \times \mathcal{X}_{\mathcal{W}^c}, \end{aligned} \quad (114)$$

$$\leq \frac{\epsilon}{|\mathcal{X}_{\mathcal{W}}|} \left(\frac{1 + |\mathcal{X}_{\mathcal{W}^c}|}{|\mathcal{X}_{\mathcal{W}^c}|} \right), \quad (115)$$

where (q) follows from (112) and (113), thus, proving the conditional strong typicality of $\vec{\mathbf{x}}_{\mathcal{W}^c}$ given $\vec{\mathbf{x}}_{\mathcal{W}}$. Note that equation (115) imposes the definition of the conditionally strongly typical set as

$$\mathcal{T}_{\epsilon}^{(n)}(\vec{\mathbf{X}}_{\mathcal{W}^c}|\vec{\mathbf{x}}_{\mathcal{W}}) = \left\{ \vec{\mathbf{x}}_{\mathcal{W}^c} : \begin{array}{ll} \left| \frac{1}{n} N(a, b|\vec{\mathbf{x}}_{\mathcal{W}}, \vec{\mathbf{x}}_{\mathcal{W}^c}) - p_{\mathbf{X}_{\mathcal{W}^c}|\mathbf{X}_{\mathcal{W}}}^i(b|a) N(a|\vec{\mathbf{x}}_{\mathcal{W}}) \right| \leq \frac{\epsilon(1+|\mathcal{X}_{\mathcal{W}^c}|)}{|\mathcal{X}_{\mathcal{W}}||\mathcal{X}_{\mathcal{W}^c}|} & \text{if } p_{\mathbf{X}_{\mathcal{W}^c}|\mathbf{X}_{\mathcal{W}}}^i(b|a) > 0 \\ N(a, b|\vec{\mathbf{x}}_{\mathcal{W}}, \vec{\mathbf{x}}_{\mathcal{W}^c}) = 0, & \text{if } p_{\mathbf{X}_{\mathcal{W}^c}|\mathbf{X}_{\mathcal{W}}}^i(b|a) = 0 \end{array} \right\}. \quad (116)$$

■

Lemma 4. *The number of codewords $\vec{\mathbf{x}}_{\mathcal{W}^c}$ of the conditional type $Q_{\vec{\mathbf{x}}_{\mathcal{W}^c}|\vec{\mathbf{x}}_{\mathcal{W}}}$ (w.r.t. a codeword $\vec{\mathbf{x}}_{\mathcal{W}}$), which are fallen in a bin $\mathcal{B}(\vec{\mathbf{m}})$ corresponding to the transmitted message $\vec{\mathbf{m}}$ is tightly concentrated around its expected value with high probability, i.e.,*

$$\left| Q_{\vec{\mathbf{x}}_{\mathcal{W}^c}|\vec{\mathbf{x}}_{\mathcal{W}}} \cap \mathcal{B}(\vec{\mathbf{m}}) \right| \rightarrow \mathbb{E}_{\mathcal{B}} \left[\left| Q_{\vec{\mathbf{x}}_{\mathcal{W}^c}|\vec{\mathbf{x}}_{\mathcal{W}}} \cap \mathcal{B}(\vec{\mathbf{m}}) \right| \right] = \frac{|Q_{\vec{\mathbf{x}}_{\mathcal{W}^c}|\vec{\mathbf{x}}_{\mathcal{W}}|}{|\vec{\mathbf{M}}|} \quad w.h.p. \quad (117)$$

provided

$$|\vec{\mathbf{M}}| \leq 2^n \left(H(p_{\mathbf{X}_{\mathcal{W}^c}|\mathbf{X}_{\mathcal{W}}}^i(\cdot|\cdot)) - \tilde{r}_d^1 \right), \quad (118)$$

where

$$\tilde{r}_d^1 \geq 2\epsilon \log_2 \left(\frac{2^C}{\epsilon} \right) + \frac{2^C \log_2(n+1)}{n}. \quad (119)$$

Proof: Recall that the conditional type of a codeword sequence $\vec{\mathbf{x}}_{\mathcal{W}^c}$ given $\vec{\mathbf{x}}_{\mathcal{W}}$ is a stochastic matrix that gives the proportion of times a particular symbol of $\mathcal{X}_{\mathcal{W}^c}$ occurred with each symbol of $\mathcal{X}_{\mathcal{W}}$ in the pair $[\vec{\mathbf{x}}_{\mathcal{W}}; \vec{\mathbf{x}}_{\mathcal{W}^c}]$. In particular, for $(a, b) \in \mathcal{X}_{\mathcal{W}} \times \mathcal{X}_{\mathcal{W}^c}$, the conditional type is defined as $Q_{\vec{\mathbf{x}}_{\mathcal{W}^c}|\vec{\mathbf{x}}_{\mathcal{W}}}(b|a) = \frac{N(a, b|\vec{\mathbf{x}}_{\mathcal{W}}, \vec{\mathbf{x}}_{\mathcal{W}^c})}{N(a|\vec{\mathbf{x}}_{\mathcal{W}})}$.

Now, by Chernoff bound, we have

$$\begin{aligned} \Pr_{\mathcal{B}} \left(\left| |Q_{\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}}} \cap \mathcal{B}(\vec{\mathbf{m}})| - \mathbb{E}_{\mathcal{B}} [|Q_{\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}}} \cap \mathcal{B}(\vec{\mathbf{m}})|] \right| \geq \epsilon \mathbb{E}_{\mathcal{B}} [|Q_{\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}}} \cap \mathcal{B}(\vec{\mathbf{m}})|] \right) \\ \leq 2 \exp \left(\frac{-\epsilon^2 \mathbb{E}_{\mathcal{B}} [|Q_{\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}}} \cap \mathcal{B}(\vec{\mathbf{m}})|]}{3} \right). \end{aligned} \quad (120)$$

Note that the expected number of codeword (balls) of a particular conditional type that fall into any message bin is

$$\mathbb{E}_{\mathcal{B}} [|Q_{\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}}}(b|a) \cap \mathcal{B}(\vec{\mathbf{m}})|] = \frac{|Q_{\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}}}(b|a)|}{|\vec{\mathbf{M}}|}, \quad (121)$$

for the total number of message bins is equal to the total number of messages.

Next, to upper bound (120), let us find a lower bound on the cardinality of the conditional type class. If we multiply the stochastic matrix associated with the conditional type $Q_{\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}}}$ by the type of $\vec{x}_{\mathcal{W}}$, denoted as $Q_{\vec{x}_{\mathcal{W}}}$, we get the joint type of $\vec{x} = [\vec{x}_{\mathcal{W}}; \vec{x}_{\mathcal{W}^c}]$ denoted as $Q_{\vec{x}}$. Now, the conditional type class can be considered as the set $Q_{\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}}} = \{\vec{x}_{\mathcal{W}^c} : [\vec{x}_{\mathcal{W}}; \vec{x}_{\mathcal{W}^c}] \in Q_{\vec{x}}\}$. Observe that $|Q_{\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}}|$ is constant for a given $\vec{x}_{\mathcal{W}} \in Q_{\vec{x}_{\mathcal{W}}}$, and thus, $|Q_{\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}}| = \frac{|Q_{\vec{x}}|}{|Q_{\vec{x}_{\mathcal{W}}}|}$. Then, by lower bounding $|Q_{\vec{x}}|$ and upper bounding $|Q_{\vec{x}_{\mathcal{W}}|$ using [13], we have

$$\frac{1}{(n+1)^{|\mathcal{X}_{\mathcal{W}}||\mathcal{X}_{\mathcal{W}^c}|}} \frac{2^{nH(Q_{\vec{x}})}}{2^{nH(Q_{\vec{x}_{\mathcal{W}}})}} \leq |Q_{\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}}| \quad (122)$$

Therefore, if $|\vec{\mathbf{M}}| \leq 2^{n(H(p_{\mathbf{X}_{\mathcal{W}^c}|\mathbf{X}_{\mathcal{W}}}^i(\cdot|\cdot)) - \tilde{r}_d^1)}$, using (121) and (122), the probability in (120) can be upper bounded by

$$\begin{aligned} \Pr_{\mathcal{B}} \left(\left| |Q_{\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}}} \cap \mathcal{B}(\vec{\mathbf{m}})| - \mathbb{E}_{\mathcal{B}} [|Q_{\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}}} \cap \mathcal{B}(\vec{\mathbf{m}})|] \right| \geq \epsilon \mathbb{E}_{\mathcal{B}} [|Q_{\vec{x}_{\mathcal{W}^c}|\vec{x}_{\mathcal{W}}} \cap \mathcal{B}(\vec{\mathbf{m}})|] \right) \\ \leq 2 \exp \left(\frac{-\epsilon^2 2^{n[H(Q_{\vec{x}}) - H(Q_{\vec{x}_{\mathcal{W}}}) - H(p_{\mathbf{X}_{\mathcal{W}^c}|\mathbf{X}_{\mathcal{W}}}^i(\cdot|\cdot)) - \frac{|\mathcal{X}_{\mathcal{W}}||\mathcal{X}_{\mathcal{W}^c}| \log_2(n+1)}{n} + \tilde{r}_d^1]}}{3} \right). \end{aligned} \quad (123)$$

This probability will be close to zero if

$$\left[H(Q_{\vec{x}}) - H(Q_{\vec{x}_{\mathcal{W}}}) - H(p_{\mathbf{X}_{\mathcal{W}^c}|\mathbf{X}_{\mathcal{W}}}^i(\cdot|\cdot)) - \frac{|\mathcal{X}_{\mathcal{W}}||\mathcal{X}_{\mathcal{W}^c}| \log_2(n+1)}{n} + \tilde{r}_d^1 \right] > 0. \quad (124)$$

The next task is to characterize $\tilde{r}_d^1$ such that the aforementioned inequality holds. For that, let $\theta(Q_{\vec{x}})$ denote the variation distance between the type $Q_{\vec{x}}$ and $p_{\mathbf{X}}^i(\cdot)$. Then, since $\vec{x}$ is strongly typical, it is straightforward to show (using similar arguments as in Lemma 3) that $\theta(Q_{\vec{x}}) \leq \frac{\epsilon}{2}$. Similarly, let $\theta(Q_{\vec{x}_{\mathcal{W}}})$ denote the variation distance between the type $Q_{\vec{x}_{\mathcal{W}}}$ and $p_{\mathbf{X}_{\mathcal{W}}}^i(\cdot)$. Then, using the strong typicality of $\vec{x}_{\mathcal{W}}$, we can show that $\theta(Q_{\vec{x}_{\mathcal{W}}}) \leq \frac{\epsilon}{2}$.

Now, observe that

$$|H(Q_{\vec{x}}) - H(Q_{\vec{x}_{\mathcal{W}}}) - H(p_{\mathbf{X}_{\mathcal{W}^c}|\mathbf{X}_{\mathcal{W}}}^i(\cdot|\cdot))| \stackrel{(a)}{=} |H(Q_{\vec{x}}) - H(Q_{\vec{x}_{\mathcal{W}}}) - H(p_{\mathbf{X}}^i(\cdot)) + H(p_{\mathbf{X}_{\mathcal{W}}}^i(\cdot))| \quad (125)$$

$$\leq |H(Q_{\vec{x}}) - H(p_{\mathbf{X}}^i(\cdot))| + |H(p_{\mathbf{X}_{\mathcal{W}}}^i(\cdot)) - H(Q_{\vec{x}_{\mathcal{W}}})| \quad (126)$$

$$\stackrel{(b)}{\leq} 2\theta(Q_{\vec{x}}) \log_2 \frac{|\mathcal{X}|}{2\theta(Q_{\vec{x}})} + 2\theta(Q_{\vec{x}_{\mathcal{W}}}) \log_2 \frac{|\mathcal{X}_{\mathcal{W}}|}{2\theta(Q_{\vec{x}_{\mathcal{W}}})} \quad (127)$$

$$\stackrel{(c)}{\leq} \epsilon \log_2 \frac{|\mathcal{X}|}{\epsilon} + \epsilon \log_2 \frac{|\mathcal{X}_{\mathcal{W}}|}{\epsilon}, \quad (128)$$

where, to obtain (a), we use the fact that the distribution $p_{\mathbf{X}}^i(\cdot)$ can be written as $p_{\mathbf{X}}^i(c) = p_{\mathbf{X}_{\mathcal{W}}, \mathbf{X}_{\mathcal{W}^c}}^i(a, b) = p_{\mathbf{X}_{\mathcal{W}^c}|\mathbf{X}_{\mathcal{W}}}^i(b|a) p_{\mathbf{X}_{\mathcal{W}}}^i(a)$ for $c \in \mathcal{X}$, $(a, b) \in \mathcal{X}_{\mathcal{W}} \times \mathcal{X}_{\mathcal{W}^c}$; (b) follows from the result [12] (Lemma 2.7):

$|H(Q) - H(p_{\mathbf{X}}^i(\cdot))| \leq 2\theta(Q) \log_2 \frac{|\mathcal{X}|}{2\theta(Q)}$ for $0 \leq \theta(Q) \leq \frac{1}{4}$; and (c) follows since $-\log_2 t$ is an increasing function of t for $0 \leq t \leq \frac{1}{2}$.

Therefore, LHS in equation (124) becomes

$$\begin{aligned} H(Q_{\vec{\mathbf{x}}}) - H(Q_{\vec{\mathbf{x}}_{\mathcal{W}}}) - H(p_{\mathbf{X}_{\mathcal{W}^c}|\mathbf{X}_{\mathcal{W}}}^i(\cdot|\cdot)) - \frac{|\mathcal{X}_{\mathcal{W}}||\mathcal{X}_{\mathcal{W}^c}|\log_2(n+1)}{n} + \tilde{r}_d^1 \\ \geq 2\epsilon \log_2 \frac{\epsilon}{2^C} - \frac{2^C \log_2(n+1)}{n} + \tilde{r}_d^1, \end{aligned} \quad (129)$$

since $|\mathcal{X}_{\mathcal{W}}||\mathcal{X}_{\mathcal{W}^c}| = |\mathcal{X}| = 2^C$. Above term (129) is positive if $\tilde{r}_d^1 \geq 2\epsilon \log_2 \left(\frac{2^C}{\epsilon} \right) + \frac{2^C \log_2(n+1)}{n}$, and in that case, $|Q_{\vec{\mathbf{x}}_{\mathcal{W}^c}|\vec{\mathbf{x}}_{\mathcal{W}}} \cap \mathcal{B}(\vec{\mathbf{m}})| \rightarrow \mathbb{E}_{\mathcal{B}} [|Q_{\vec{\mathbf{x}}_{\mathcal{W}^c}|\vec{\mathbf{x}}_{\mathcal{W}}} \cap \mathcal{B}(\vec{\mathbf{m}})|] = \frac{|Q_{\vec{\mathbf{x}}_{\mathcal{W}^c}|\vec{\mathbf{x}}_{\mathcal{W}}|}{|\mathcal{M}|}$ with high probability. ■

APPENDIX E MODIFICATION FOR ACHIEVING RELIABILITY

First, Alice bins uniformly at random the codewords from the ϵ -strongly typical set (w.r.t. scalar active distribution $p_{\mathbf{X}}^a(\cdot)$) as mentioned in Procedure 1. Next, when Alice is active, she transmits the covert messages at the rate $R_d = R - \frac{\log_2 n}{n}$. (For simplicity, we assume that n is some power of 2.) Therefore, essentially Alice considers some $\frac{2^{nR}}{n}$ message bins; without loss of generality assume that the first $\frac{2^{nR}}{n}$ bins are selected. Denote the set of bins as $\mathcal{M}_d = \{1, \dots, 2^{nR - \log_2 n}\}$. Notice that Alice's active codebook is

$$\mathcal{C} = \cup_{\vec{\mathbf{m}} \in \mathcal{M}_d} \mathcal{B}(\vec{\mathbf{m}}).$$

To prove reliability, note that when Alice is innocent she generates $\vec{\mathbf{x}}$ according to $p_{\mathbf{X}}^i(\cdot)$. Bob will wrongly estimate Alice to be active if her codeword falls in one of the first $\frac{2^{nR}}{n}$ message bins. Because of random binning, the probability of this event is $\frac{1}{n}$. When Alice is active, Bob can always decode the correct covert message since the encoding is uniquely decidable.

Now, let us prove that this scheme continues to be deniable. For that, consider the marginal innocent distribution $p_{\vec{\mathbf{x}}_{\mathcal{W}}}^i(\vec{\mathbf{x}}_{\mathcal{W}})$.

$$p_{\vec{\mathbf{x}}_{\mathcal{W}}}^i(\vec{\mathbf{x}}_{\mathcal{W}}) \stackrel{(a)}{=} \sum_{\vec{\mathbf{x}}'_{\mathcal{W}^c}} p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}}'_{\mathcal{W}^c}, \vec{\mathbf{x}}_{\mathcal{W}}) \quad (130)$$

$$= \sum_{\vec{\mathbf{x}}': \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}}} p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}}') \quad (131)$$

$$\stackrel{(b)}{=} \sum_{\substack{\vec{\mathbf{x}}' \in \mathcal{T}_{\epsilon}^{(n)}(\vec{\mathbf{x}}) \\ \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}}}} p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}}') + \sum_{\substack{\vec{\mathbf{x}}' \notin \mathcal{T}_{\epsilon}^{(n)}(\vec{\mathbf{x}}) \\ \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}}}} p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}}') \quad (132)$$

$$\stackrel{(c)}{=} \sum_{\vec{\mathbf{x}}': \vec{\mathbf{x}}'_{\mathcal{W}^c} \in \mathcal{T}_{\epsilon}^{(n)}(\vec{\mathbf{x}}_{\mathcal{W}^c}|\vec{\mathbf{x}}_{\mathcal{W}})} p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}}') + \sum_{\vec{\mathbf{x}}': \vec{\mathbf{x}}'_{\mathcal{W}^c} \notin \mathcal{T}_{\epsilon}^{(n)}(\vec{\mathbf{x}}_{\mathcal{W}^c}|\vec{\mathbf{x}}_{\mathcal{W}})} p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}}'), \quad (133)$$

where (a) follows from the definition of the marginal distribution, (b) follows by splitting the summation, and (c) is due to the following reason: we proved in Lemma 3 in appendix D that, for any strongly typical $\vec{\mathbf{x}}$ (w.r.t. $p_{\mathbf{X}}^i(\cdot)$), the sequence $\vec{\mathbf{x}}_{\mathcal{W}}$ observed by Willie is also strongly typical (w.r.t. $p_{\mathbf{X}_{\mathcal{W}}}^i(\cdot)$). Furthermore, the sequence not observed by Willie, $\vec{\mathbf{x}}_{\mathcal{W}^c}$, is conditionally strongly typical given $\vec{\mathbf{x}}_{\mathcal{W}}$ (Lemma 3). Thus, the set $\{\vec{\mathbf{x}}' : \vec{\mathbf{x}}' \in \mathcal{T}_{\epsilon}^{(n)}, \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}}\}$ is the same as the conditionally strongly typical set $\mathcal{T}_{\epsilon}^{(n)}(\vec{\mathbf{x}}_{\mathcal{W}^c}|\vec{\mathbf{x}}_{\mathcal{W}})$ (w.r.t. $p_{\mathbf{X}_{\mathcal{W}^c}|\mathbf{X}_{\mathcal{W}}}^i(\cdot|\cdot)$).

Now, consider the marginal induced distribution as follow.

$$\hat{p}_{\vec{\mathbf{x}}_{\mathcal{W}}}(\vec{\mathbf{x}}_{\mathcal{W}}) = \sum_{\vec{\mathbf{x}}'_{\mathcal{W}^c}} \hat{p}_{\vec{\mathbf{x}}}(\vec{\mathbf{x}}'_{\mathcal{W}^c}, \vec{\mathbf{x}}_{\mathcal{W}}) = \sum_{\vec{\mathbf{x}}': \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}}} \hat{p}_{\vec{\mathbf{x}}}(\vec{\mathbf{x}}') \quad (134)$$

$$\stackrel{(d)}{=} \sum_{\vec{\mathbf{m}} \in \mathcal{M}_d} \sum_{\vec{\mathbf{x}}': \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}}} \Pr_{\vec{\mathbf{M}}, \text{Enc}}(\vec{\mathbf{x}}', \vec{\mathbf{m}} | \mathbf{T} = 1) \quad (135)$$

$$= \sum_{\vec{\mathbf{m}} \in \mathcal{M}_d} \sum_{\vec{\mathbf{x}}': \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}}} \Pr_{\vec{\mathbf{M}}, \text{Enc}}(\vec{\mathbf{x}}' | \vec{\mathbf{m}}, \mathbf{T} = 1) \Pr_{\vec{\mathbf{M}}}(\vec{\mathbf{m}} | \mathbf{T} = 1) \quad (136)$$

$$\stackrel{(e)}{=} \sum_{\vec{\mathbf{m}} \in \mathcal{M}_d} \sum_{\substack{\vec{\mathbf{x}}': \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}} \\ \vec{\mathbf{x}}' \in \mathcal{B}(\vec{\mathbf{m}})}} \frac{p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}}')}{\Pr_{\mathcal{B}}(\mathcal{B}(\vec{\mathbf{m}}))} \frac{1}{2^{nR_d}}, \quad (137)$$

where (d) follows from the law of total probability, and (e) follows from the definition of the stochastic encoding and that the messages are uniform over 2^{nR_d} choices.

We have proved in Lemma 1 that, with high probability,

$$(1 - \epsilon_1) \frac{1}{2^{nR}} \leq \Pr_{\mathcal{B}}(\mathcal{B}(\vec{\mathbf{m}})) \leq (1 + \epsilon_1) \frac{1}{2^{nR}}.$$

Therefore, substituting in (137), we get

$$\frac{2^{n(R-R_d)}}{1 + \epsilon_1} \sum_{\vec{\mathbf{m}} \in \mathcal{M}_d} \sum_{\substack{\vec{\mathbf{x}}': \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}} \\ \vec{\mathbf{x}}' \in \mathcal{B}(\vec{\mathbf{m}})}} p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}}') \leq \hat{p}_{\vec{\mathbf{x}}_{\mathcal{W}}}(\vec{\mathbf{x}}_{\mathcal{W}}) \leq \frac{2^{n(R-R_d)}}{1 - \epsilon_1} \sum_{\vec{\mathbf{m}} \in \mathcal{M}_d} \sum_{\substack{\vec{\mathbf{x}}': \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}} \\ \vec{\mathbf{x}}' \in \mathcal{B}(\vec{\mathbf{m}})}} p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}}'). \quad (138)$$

Note that

$$\sum_{\vec{\mathbf{m}} \in \mathcal{M}_d} \sum_{\substack{\vec{\mathbf{x}}': \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}} \\ \vec{\mathbf{x}}' \in \mathcal{B}(\vec{\mathbf{m}})}} p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}}') \stackrel{(f)}{=} \sum_{\substack{\vec{\mathbf{x}}': \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}} \\ \vec{\mathbf{x}}' \in \cup_{\vec{\mathbf{m}} \in \mathcal{M}_d} \mathcal{B}(\vec{\mathbf{m}})}} p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}}') \stackrel{(g)}{=} \sum_{\substack{\vec{\mathbf{x}}': \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}} \\ \vec{\mathbf{x}}' \in \mathcal{T}_{\epsilon}^{(n)}(\vec{\mathbf{X}}) \cap \mathcal{C}}} p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}}'),$$

where (f) follows because the sets $\mathcal{B}(\vec{\mathbf{m}})$ are disjoint for all $\vec{\mathbf{m}}$, and (g) is due to $\cup_{\vec{\mathbf{m}} \in \mathcal{M}_d} \mathcal{B}(\vec{\mathbf{m}}) = \mathcal{C} = \mathcal{T}_{\epsilon}^{(n)}(\vec{\mathbf{X}}) \cap \mathcal{C}$. Again, following the reasoning of (c) above, we have that the set $\{\vec{\mathbf{x}}' : \vec{\mathbf{x}}' \in \mathcal{T}_{\epsilon}^{(n)}, \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}}\}$ is the same as the conditionally strongly typical set $\mathcal{T}_{\epsilon}^{(n)}(\vec{\mathbf{X}}_{\mathcal{W}^c} | \vec{\mathbf{x}}_{\mathcal{W}})$ (w.r.t. $p_{\vec{\mathbf{x}}_{\mathcal{W}^c} | \vec{\mathbf{x}}_{\mathcal{W}}}^i(\cdot | \cdot)$).

Now, since we can decompose the conditional strongly typical set into number of type classes, we can write the term above as

$$\sum_{\substack{\vec{\mathbf{x}}': \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}} \\ \vec{\mathbf{x}}' \in \mathcal{T}_{\epsilon}^{(n)}(\vec{\mathbf{X}}) \cap \mathcal{C}}} p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}}') = \sum_{\vec{\mathbf{x}}': \vec{\mathbf{x}}'_{\mathcal{W}^c} \in \mathcal{T}_{\epsilon}^{(n)}(\vec{\mathbf{X}}_{\mathcal{W}^c} | \vec{\mathbf{x}}_{\mathcal{W}}) \cap \mathcal{C}} p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}}') \quad (139)$$

$$= \sum_{Q \in \mathcal{T}_{\epsilon}^{(n)}(\vec{\mathbf{X}}_{\mathcal{W}^c} | \vec{\mathbf{x}}_{\mathcal{W}})} \sum_{\vec{\mathbf{x}}': \vec{\mathbf{x}}'_{\mathcal{W}^c} \in Q \cap \mathcal{C}} p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}}') \quad (140)$$

$$= \sum_{Q \in \mathcal{T}_{\epsilon}^{(n)}(\vec{\mathbf{X}}_{\mathcal{W}^c} | \vec{\mathbf{x}}_{\mathcal{W}})} |Q \cap \mathcal{C}| p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}}'_Q) \quad (141)$$

$$\stackrel{(h)}{=} \sum_{Q \in \mathcal{T}_{\epsilon}^{(n)}(\vec{\mathbf{X}}_{\mathcal{W}^c} | \vec{\mathbf{x}}_{\mathcal{W}})} \frac{|Q|}{n} p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}}'_Q) \quad \text{w.h.p.} \quad (142)$$

$$\stackrel{(i)}{=} \frac{1}{n} \sum_{\substack{\vec{\mathbf{x}}': \vec{\mathbf{x}}'_{\mathcal{W}} = \vec{\mathbf{x}}_{\mathcal{W}} \\ \vec{\mathbf{x}}' \in \mathcal{T}_{\epsilon}^{(n)}(\vec{\mathbf{X}})}} p_{\vec{\mathbf{x}}}^i(\vec{\mathbf{x}}') \quad \text{w.h.p.} \quad (143)$$

where (h) follows from the Chernoff bound, and (i) follows due to $\left\{ \vec{x}' : \vec{x}' \in \mathcal{T}_\epsilon^{(n)}, \vec{x}'_{\mathcal{W}} = \vec{x}_{\mathcal{W}} \right\} = \mathcal{T}_\epsilon^{(n)} \left(\vec{\mathbf{X}}_{\mathcal{W}^c} | \vec{x}_{\mathcal{W}} \right)$.

Therefore, with high probability, we have

$$\frac{2^{n(R-R_d)}}{n(1+\epsilon_1)} \sum_{\substack{\vec{x}': \vec{x}'_{\mathcal{W}} = \vec{x}_{\mathcal{W}} \\ \vec{x}' \in \mathcal{T}_\epsilon^{(n)}(\vec{\mathbf{X}})}} p_{\vec{\mathbf{X}}}^i(\vec{x}') \leq \hat{p}_{\vec{\mathbf{X}}_{\mathcal{W}}}(\vec{x}_{\mathcal{W}}) \leq \frac{2^{n(R-R_d)}}{n(1-\epsilon_1)} \sum_{\substack{\vec{x}': \vec{x}'_{\mathcal{W}} = \vec{x}_{\mathcal{W}} \\ \vec{x}' \in \mathcal{T}_\epsilon^{(n)}(\vec{\mathbf{X}})}} p_{\vec{\mathbf{X}}}^i(\vec{x}'). \quad (144)$$

Since $2^{n(R-R_d)} = n$, we have

$$(1-\epsilon_1)\hat{p}_{\vec{\mathbf{X}}_{\mathcal{W}}}(\vec{x}_{\mathcal{W}}) \leq \sum_{\vec{x}': \vec{x}'_{\mathcal{W}^c} \in \mathcal{T}_\epsilon^{(n)}(\vec{\mathbf{X}}_{\mathcal{W}^c} | \vec{x}_{\mathcal{W}})} p_{\vec{\mathbf{X}}}^i(\vec{x}') \leq (1+\epsilon_1)\hat{p}_{\vec{\mathbf{X}}_{\mathcal{W}}}(\vec{x}_{\mathcal{W}}). \quad (145)$$

Thus,

$$\left| \hat{p}_{\vec{\mathbf{X}}_{\mathcal{W}}}(\vec{x}_{\mathcal{W}}) - \sum_{\vec{x}': \vec{x}'_{\mathcal{W}^c} \in \mathcal{T}_\epsilon^{(n)}(\vec{\mathbf{X}}_{\mathcal{W}^c} | \vec{x}_{\mathcal{W}})} p_{\vec{\mathbf{X}}}^i(\vec{x}') \right| \leq \epsilon_1 \hat{p}_{\vec{\mathbf{X}}_{\mathcal{W}}}(\vec{x}_{\mathcal{W}}). \quad (146)$$

Now, recall that (from (133)) we have,

$$p_{\vec{\mathbf{X}}_{\mathcal{W}}}^i(\vec{x}_{\mathcal{W}}) = \underbrace{\sum_{\vec{x}': \vec{x}'_{\mathcal{W}^c} \in \mathcal{T}_\epsilon^{(n)}(\vec{\mathbf{X}}_{\mathcal{W}^c} | \vec{x}_{\mathcal{W}})} p_{\vec{\mathbf{X}}}^i(\vec{x}')}_1 + \underbrace{\sum_{\vec{x}': \vec{x}'_{\mathcal{W}^c} \notin \mathcal{T}_\epsilon^{(n)}(\vec{\mathbf{X}}_{\mathcal{W}^c} | \vec{x}_{\mathcal{W}})} p_{\vec{\mathbf{X}}}^i(\vec{x}')}_2.$$

Now, consider the total variation distance between the marginals

$$TVD = \frac{1}{2} \sum_{\vec{x}_{\mathcal{W}}} \left| \hat{p}_{\vec{\mathbf{X}}_{\mathcal{W}}}(\vec{x}_{\mathcal{W}}) - p_{\vec{\mathbf{X}}_{\mathcal{W}}}^i(\vec{x}_{\mathcal{W}}) \right| \quad (147)$$

$$\stackrel{(j)}{=} \frac{1}{2} \sum_{\vec{x}_{\mathcal{W}}} \left| \hat{p}_{\vec{\mathbf{X}}_{\mathcal{W}}}(\vec{x}_{\mathcal{W}}) - \sum_{\vec{x}': \vec{x}'_{\mathcal{W}^c} \in \mathcal{T}_\epsilon^{(n)}(\vec{\mathbf{X}}_{\mathcal{W}^c} | \vec{x}_{\mathcal{W}})} p_{\vec{\mathbf{X}}}^i(\vec{x}') - \sum_{\vec{x}': \vec{x}'_{\mathcal{W}^c} \notin \mathcal{T}_\epsilon^{(n)}(\vec{\mathbf{X}}_{\mathcal{W}^c} | \vec{x}_{\mathcal{W}})} p_{\vec{\mathbf{X}}}^i(\vec{x}') \right| \quad (148)$$

$$\leq \frac{1}{2} \sum_{\vec{x}_{\mathcal{W}}} \left| \hat{p}_{\vec{\mathbf{X}}_{\mathcal{W}}}(\vec{x}_{\mathcal{W}}) - \sum_{\vec{x}': \vec{x}'_{\mathcal{W}^c} \in \mathcal{T}_\epsilon^{(n)}(\vec{\mathbf{X}}_{\mathcal{W}^c} | \vec{x}_{\mathcal{W}})} p_{\vec{\mathbf{X}}}^i(\vec{x}') \right| + \frac{1}{2} \sum_{\vec{x}_{\mathcal{W}}} \left| \sum_{\vec{x}': \vec{x}'_{\mathcal{W}^c} \notin \mathcal{T}_\epsilon^{(n)}(\vec{\mathbf{X}}_{\mathcal{W}^c} | \vec{x}_{\mathcal{W}})} p_{\vec{\mathbf{X}}}^i(\vec{x}') \right| \quad (149)$$

$$\stackrel{(k)}{\leq} \frac{1}{2} \sum_{\vec{x}_{\mathcal{W}}} \epsilon_1 \hat{p}_{\vec{\mathbf{X}}_{\mathcal{W}}}(\vec{x}_{\mathcal{W}}) + \frac{1}{2} \sum_{\vec{x}_{\mathcal{W}}} \left| \sum_{\vec{x}': \vec{x}'_{\mathcal{W}^c} \notin \mathcal{T}_\epsilon^{(n)}(\vec{\mathbf{X}}_{\mathcal{W}^c} | \vec{x}_{\mathcal{W}})} p_{\vec{\mathbf{X}}}^i(\vec{x}') \right| \quad (150)$$

$$\stackrel{(l)}{\leq} \frac{1}{2} \epsilon_1 + \sum_{\vec{x}' \notin \mathcal{T}_\epsilon^{(n)}(\vec{\mathbf{X}})} p_{\vec{\mathbf{X}}}^i(\vec{x}') \quad (151)$$

$$\stackrel{(m)}{\leq} \frac{1}{2} \epsilon_1 + \frac{1}{2} 2 |\mathcal{X}| \exp(-2\epsilon^2 n), \quad (152)$$

where (j) is due to (133), (k) is from (146), (l) is due to $\cup_{\vec{x}_{\mathcal{W}}} \mathcal{T}_\epsilon^{(n)}(\vec{\mathbf{X}}_{\mathcal{W}^c} | \vec{x}_{\mathcal{W}}) = \mathcal{T}_\epsilon^{(n)}(\vec{\mathbf{X}})$, and (m) is due to typicality.

ACKNOWLEDGMENT

Author Swanand Kadhe would like to thank Dr. Chung Chan for helpful discussions.

REFERENCES

- [1] U. Maurer, "Secret key agreement by public discussion from common information," *Transactions on Information Theory, IEEE*, vol. 39, no. 3, pp. 733–742, 1993.
- [2] U. Maurer and S. Wolf, "Information-theoretic key agreement: From weak to strong secrecy for free," *Advances in Cryptology – EUROCRYPT*, pp. 351–368, 2000.
- [3] D. McCoy, K. Bauer, D. Grunwald, T. Kohno, and D. Sicker, "Shining light in dark places: Understanding the tor network," *Privacy Enhancing Technologies*, pp. 63–76, 2008.
- [4] H. Sousa-Pinto, D. E. Lucani, and J. Barros, "Hide and code: Session anonymity in wireless line networks with coded packets," in *Proceedings of the Information Theory and Applications Workshop (ITA)*. IEEE, 2012, pp. 262–268.
- [5] C. Cachin, "An information-theoretic model for steganography," in *Information Hiding*, 1998, pp. 306–318.
- [6] Y. Wang and P. Moulin, "Perfectly secure steganography: Capacity, error exponents, and code constructions," *IEEE Transactions on Information Theory*, vol. 54, no. 6, pp. 2706–2722, 2008.
- [7] B. Ryabko and D. Ryabko, "Asymptotically optimal perfect steganographic systems," *Problems of Information Transmission*, vol. 45, no. 2, pp. 184–190, 2009.
- [8] B. Bash, D. Goeckel, and D. Towsley, "Square root law for communication with low probability of detection on AWGN channels," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, 2012, pp. 448–452.
- [9] P. H. Che, M. Bakshi, and S. Jaggi, "Reliable deniable communication: Hiding messages in noise," in *Proceedings of ISIT*. IEEE, 2013.
- [10] J. Hou and G. Kramer, "Effective secrecy: Reliability, confusion and stealth," *CoRR*, vol. abs/1311.1411, 2013.
- [11] E. Lehmann and J. Romano, "Testing statistical hypotheses, (texts in statistics)," 2005.
- [12] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems, 2nd ed.* Academic Press, 2011.
- [13] T. M. Cover and J. A. Thomas, *Elements of information theory*. Wiley-interscience, 2012.