Square Root Law for Communication with Low Probability of Detection on AWGN Channels

Boulat A. Bash, Student Member, IEEE, Dennis Goeckel, Fellow, IEEE,
and Don Towsley, Fellow, IEEE

Abstract

We present a square root limit on low probability of detection (LPD) communication over additive white Gaussian noise (AWGN) channels. Specifically, if a warden has an AWGN channel to the transmitter with non-zero noise power, we prove that $o(\sqrt{n})$ bits can be sent from the transmitter to the receiver in $n$ AWGN channel uses with probability of detection by the warden less than $\epsilon$ for any $\epsilon > 0$. Moreover, in most practical scenarios, a lower bound on the noise power on the warden’s channel to the transmitter is known and $O(\sqrt{n})$ bits can be covertly sent in $n$ channel uses. Conversely, attempting to transmit more than $O(\sqrt{n})$ bits either results in detection by the warden with probability one or a non-zero probability of decoding error as $n \to \infty$. Further, we show that LPD communication on the AWGN channel allows one to send a non-zero symbol on every channel use, in contrast to what might be expected from the square root law found recently in image-based steganography.

B. A. Bash and D. Towsley are with the Computer Science Department, University of Massachusetts, Amherst, Massachusetts. D. Goeckel is with the Electrical and Computer Engineering Department, University of Massachusetts, Amherst, Massachusetts.

This research was sponsored by the National Science Foundation under grants CNS-0905349 and CNS-1018464, and by the U.S. Army Research Laboratory and the U.K. Ministry of Defence under Agreement Number W911NF-06-3-0001. The views and conclusions contained in this document are those of the author(s) and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

An abridged version of this manuscript was submitted to the International Symposium on Information Theory (ISIT) 2012.
I. INTRODUCTION

Securing information transmitted over wireless links is of paramount concern for consumer, industrial, and military applications. The taxonomy of network security classifies secure communication into two distinct categories: low probability of intercept (LPI) communication and low probability of detection (LPD) communication [1]. In recent years, the wireless networking community has made tremendous strides in the former area, securing data transmitted in wireless networks from interception by an untrusted eavesdropper using various encryption and key exchange protocols. However, the latter area, LPD communication, which concerns the prevention of transmissions from being detected in the first place, has been relatively underexplored.

Consider a node that is trying to send data on a wireless channel to another node so that the presence of this transmission is not detected by an eavesdropping third party. There are many real-life scenarios where this is preferable to standard cryptographic security. Encrypted data arouses suspicion, and even the most theoretically robust encryption can often be defeated by a determined adversary using non-computational methods such as side-channel analysis. Thus, the study of covert communications over LPD channels is extremely important.

We examine the fundamental limitations of covert communication over wireless links. In our system Alice transmits covert data over a wireless channel to Bob, while passive eavesdropper Warden Willie attempts to decide whether what he hears is noise or covert communication. Willie is passive in that he does not actively jam Alice’s channel. However, if he detects covert communication, he can potentially shut the channel down or otherwise punish Alice. Such a scenario requires Alice to blend in with the background noise by softly “whispering” to Bob, as sending a high-power signal will surely tip Willie off to the covert transmission.

Our problem is related to the problem of imperfect steganography, but the two problems are not the same. Standard steganography considers hiding information by altering the properties of fixed-size, finite-alphabet covertext objects (such as images or software binary code) with imperfect steganography systems allowing some fixed probability of detection of hidden information.
Covertext can be considered a type of lossless finite-alphabet channel. However, the square root law recently found in this environment [2], which states that $O(\sqrt{n})$ symbols in the covertext of size $n$ may safely be modified to hide a steganographic message, is limited in its scope. The continuous-valued channel allows us to spread hidden information across every symbol used in the transmission, thus showing that a direct application of the steganographic result quickly leads to contradiction and demonstrating the distinction between the two problems. In fact, our square root law can be viewed as generalizing the square root law for imperfect steganography.

In our scenario, Alice communicates with Bob over a channel subject to additive white Gaussian noise (AWGN), while Willie attempts to detect her transmission. The channel between Willie and Alice is also subject to AWGN. Alice sends low-power covert signals to Bob that Willie attempts to classify as either noise on his channel from Alice or Alice’s signals to Bob. If the noise on the channel between Willie and Alice has non-zero power, Alice can communicate with Bob while tolerating a certain probability of detection, which she can drive down by transmitting with low enough power. Thus, Alice potentially gets non-zero mutual information across the covert channel to Bob in $n$ uses of the channel. We state our main result that limits mutual information on the covert channel between Alice and Bob using asymptotic notation where $f(n) = O(g(n))$ denotes an asymptotically tight upper bound on $f(n)$ (i.e. there exist positive constants $m$ and $n_0$ such that $0 \leq f(n) \leq mg(n)$ for all $n \geq n_0$), $f(n) = o(g(n))$ denotes an upper bound on $f(n)$ that is not asymptotically tight (i.e. for any constant $m > 0$, there exists constant $n_0 > 0$ such that $0 \leq f(n) < mg(n)$ for all $n \geq n_0$), and $f(n) = \omega(g(n))$ denotes a lower bound on $f(n)$ that is not asymptotically tight (i.e. for any constant $m > 0$, there exists constant $n_0 > 0$ such that $0 \leq mg(n) < f(n)$ for all $n \geq n_0$) [3, Ch. 3.1]:

**Theorem** (Square root law). Suppose the channel between Alice and each of Bob and Willie experiences independent additive white Gaussian noise (AWGN) with power $\sigma_b^2 > 0$ and $\sigma_w^2 > 0$, respectively, where $\sigma_b^2$ and $\sigma_w^2$ are constants. Then, for any $\epsilon > 0$ and unknown $\sigma_w^2$, Alice can send
\( o(\sqrt{n}) \) information bits to Bob in \( n \) channel uses while maintaining a probability of detection of Alice’s transmission by Willie of less than \( \epsilon \). Moreover, if Alice can lower-bound \( \sigma_{w}^{2} \geq \hat{\sigma}_{w}^{2} \), she can send \( \mathcal{O}(\sqrt{n}) \) bits in \( n \) channel uses while maintaining a probability of detection of less than \( \epsilon \). Conversely, if Alice attempts to transmit \( \omega(\sqrt{n}) \) bits in \( n \) channel uses, then, as \( n \to \infty \), either Willie detects her with arbitrary low probability of error or Bob cannot decode her message reliably (i.e. with arbitrary low probability of decoding error).

After introducing our discrete-time channel model and hypothesis testing background in Section II, we prove the achievability of the square root law in Section III. We then prove the converse in Section IV. We discuss the mapping to the continuous-time channel and the implications of channel fading on our results, as well as the relationship to previous work in Section V, and conclude in Section VI.

### II. Prerequisites

![System framework: Alice encodes information into a vector of real symbols \( f = \{f_i\}_{i=1}^{n} \) and transmits it on an AWGN channel to Bob, while Willie attempts to classify his vector of observations of the channel from Alice \( y_w \) as either an AWGN vector \( \{z_i\}_{i=1}^{n} \) or a vector \( \{f_i + z_i\}_{i=1}^{n} \) of transmissions corrupted by AWGN.](image)

**Fig. 1.**

A. Channel Model

We use the discrete-time AWGN channel model with real-valued symbols (and defer discussion of the mapping to a continuous-time channel as well as a fading channel to Section V). Our
formal system framework is depicted in Figure 1. In our scenario, Alice transmits a vector of $n$ real-valued symbols $f = \{f_i\}_{i=1}^n$. Bob receives vector $y_b = \{y_i^{(b)}\}_{i=1}^n$ where $y_i^{(b)} = f_i + z_i^{(b)}$ with an independent and identically distributed (i.i.d.) $z_i^{(b)} \sim \mathcal{N}(0, \sigma_b^2)$. Willie observes vector $y_w = \{y_i^{(w)}\}_{i=1}^n$ where $y_i^{(w)} = f_i + z_i^{(w)}$, with i.i.d. $z_i^{(w)} \sim \mathcal{N}(0, \sigma_w^2)$. Willie uses statistical hypothesis tests on $y_w$ to determine whether Alice has communicated, which we discuss next.

B. Hypothesis Testing

Willie expects vector $y_w$ of $n$ channel readings to be consistent with his channel noise model. He performs a statistical hypothesis test on this vector, with the null hypothesis $H_0$ being that Alice is not covertly communicating. This corresponds to each sample $y_i^{(w)} \sim \mathcal{N}(0, \sigma_w^2)$ i.i.d. The alternate hypothesis $H_1$ is that Alice is transmitting, which corresponds to samples $y_i^{(w)}$ from a different distribution. Willie tolerates some false positives: cases when his statistical test incorrectly accuses Alice. This rejection of $H_0$ when it is true is known as the type I error (or false alarm), and, following the standard nomenclature, we denote its probability by $\alpha$ [4]. Willie’s test may also miss Alice’s covert transmissions. Acceptance of $H_0$ when it is false is known as the type II error (or miss), and we denote its probability by $\beta$. The sum $\alpha + \beta$ determines the performance of a hypothesis test [4].

III. Achievability of Square Root Law

In our scenario, Alice and Bob construct a covert communications system, with all the details known to Willie except for a secret key shared before communication. This follows “best practices” in security system design, as our system obeys Kerckhoffs’s law [5] because its security depends only on the key [6]. Since this work concerns the limits of covert communication, key size is not a constraint and we defer the study of key efficiency to future work.

Willie’s objective is to determine whether Alice transmitted covert data given the vector of observations $y_w$ of his channel from Alice. Denote the probability distribution of Willie’s channel observations when Alice does not transmit (i.e. when $H_0$ is true) as $P_0$, and the
probability distribution of the observations when Alice transmits (i.e. when \( H_1 \) is true) as \( P_1 \). To strengthen the achievability result, we assume that Alice’s channel input distribution, as well as the distribution of AWGN on the channel between Alice and Willie are known to Willie. Then \( P_0 \) and \( P_1 \) are known to Willie, and he can construct an optimal statistical hypothesis test that minimizes the sum of error probabilities \( \alpha + \beta \) [4, Ch. 13]. The following holds for such a test:

**Fact 1** (Theorem 13.1.1 in [4]). For the optimal test:

\[
\alpha + \beta = 1 - TV(P_0, P_1)
\]

where \( TV(P_0, P_1) \) is the total variation distance between \( P_0 \) and \( P_1 \) defined as follows:

**Definition 1** (Total variation distance [4]). The total variation distance between two probability measures \( P_0 \) and \( P_1 \) on a \( \sigma \)-algebra \( A \) is

\[
TV(P_0, P_1) = \sup\{|P_0(A) - P_1(A)| : A \in A\} = \frac{1}{2} \|p_0(x) - p_1(x)\|_1
\]

(1)

where \( p_0(x) \) and \( p_1(x) \) are densities\(^1\) of \( P_0 \) and \( P_1 \), respectively, and \( \|a - b\|_1 \) is the \( L^1 \) norm.

Since total variation lower-bounds the error of all hypothesis tests Willie can use, a clever choice of \( f \) allows Alice to limit Willie’s detector performance. Unfortunately, the total variation metric is unwieldy for the products of probability measures, which are used in the analysis of the vectors of observations. We thus use Pinsker’s Inequality:

**Fact 2** (Pinsker’s Inequality (Lemma 11.6.1 in [7])).

\[
\frac{1}{2} \|p_0(x) - p_1(x)\|_1^2 = \frac{1}{2} \left( \int_{-\infty}^{\infty} |p_0(x) - p_1(x)| dx \right)^2 \leq D(P_0\|P_1)
\]

where relative entropy \( D(P_0\|P_1) \) is defined as follows:

\(^1\text{In case of discrete } P_0 \text{ and } P_1, p_0(x) \text{ and } p_1(x) \text{ are p.m.f.'s. In our work, } P_0 \text{ and } P_1 \text{ are continuous.} \)
**Definition 2.** The relative entropy between two probability measures $P_0$ and $P_1$ is:

$$D(P_0 \parallel P_1) = \int_{\mathcal{X}} p_0(x) \ln \frac{p_0(x)}{p_1(x)} dx$$

(2)

where $\mathcal{X}$ is the support of $p_1(x)$.

If $P^n$ is the distribution of a sequence $\{X_i\}_{i=1}^n$ where each $X_i \sim P$ is i.i.d., then:

**Fact 3 (Relative Entropy Product).** From the chain rule for relative entropy [7, (2.67)]:

$$D(P^n_0 \parallel P^n_1) = nD(P_0 \parallel P_1)$$

Now we are ready to prove the achievability theorem under an average power constraint.

**Theorem 1.1 (Achievability).** Suppose Willie’s channel is subject to AWGN with constant power $\sigma_w^2 > 0$. Then Alice can maintain Willie’s sum of the probabilities of detection errors $\alpha + \beta \geq 1 - \epsilon$ for any $\epsilon > 0$ while covertly transmitting $o(\sqrt{n})$ bits to Bob over $n$ uses of an AWGN channel if $\sigma_w^2$ is unknown and $O(\sqrt{n})$ bits over $n$ channel uses if she can lower-bound $\sigma_w^2 \geq \hat{\sigma}_w^2$.

**Proof:** **Construction:** Alice’s channel encoder takes input in blocks of size $M$ bits and encodes them into codewords of length $n$ at the rate of $R = M/n$ bits/symbol. We employ random coding arguments and independently generate $2^{nR}$ codewords $\{c(W_k), k = 1, 2, \ldots, 2^{nR}\}$ from $\mathbb{R}^n$ for messages $W_k$, each according to $p_X(x) = \prod_{i=1}^n p_X(x_i)$, where $X \sim \mathcal{N}(0, P_f)$ and $P_f$ is defined later. The codebook is the secret key shared between Alice and Bob, and is not revealed to Willie. However, Willie knows how it is constructed, including the value of $P_f$.

The channel between Alice and Willie is corrupted by AWGN with power $\sigma_w^2$. Willie uses statistical hypothesis testing on a vector of $n$ channel readings $y_w$ to decide whether Alice transmitted. Next we show how Alice can limit the performance of Willie’s methods.

**Analysis:** Consider the case when Alice transmits codeword $c(W_k)$. Suppose that Willie employs a detector that implements an optimal hypothesis test on his $n$ channel readings. His null hypothesis $H_0$ is that Alice did not transmit and he observed noise on his channel. His
alternate hypothesis $H_1$ is that Alice transmitted and he observed Alice’s codeword corrupted by noise. By Fact 1, the sum of the probabilities of Willie’s detector’s errors is expressed by $\alpha + \beta = 1 - TV(P_0, P_1)$, where the total variation distance is between the distribution $P_0$ of $n$ noise readings that Willie expects to observe under his null hypothesis and the distribution $P_1$ of the covert codeword transmitted by Alice corrupted by noise. Alice can lower-bound the sum of the error probabilities by upper-bounding the total variation distance: $TV(P_0, P_1) \leq \epsilon$.

The realizations of noise $z_i^{(w)}$ in vector $z_w$ are zero-mean i.i.d. Gaussian random variables with variance $\sigma^2_w$, and, thus, $P_0 = P_w^n$ where $P_w = \mathcal{N}(0, \sigma^2_w)$. Recall that Willie does not know the codebook. Therefore, Willie’s probability distribution of the transmitted symbols is of zero-mean i.i.d. Gaussian random variables with variance $P_f$. Since noise is independent of the transmitted symbols, when Alice transmits, Willie observes vector $y_w$, where $y_i^{(w)} \sim \mathcal{N}(0, P_f + \sigma^2_w) = P_s$ is i.i.d., and thus, $P_1 = P_s^n$. By Pinsker’s Inequality and Fact 3:

$$TV(P_w^n, P_s^n) \leq \sqrt{\frac{1}{2} D(P_w^n \| P_s^n)} = \sqrt{\frac{n}{2} D(P_w \| P_s)}$$

The relative entropy follows as:

$$D(P_w \| P_s) = \int_{-\infty}^{\infty} \frac{e^{-\frac{x^2}{2\sigma_w^2}}}{\sqrt{2\pi\sigma_w}} \ln \frac{e^{-\frac{x^2}{2\sigma_w^2}} / \sqrt{2\pi\sigma_w^2}}{e^{-\frac{(x+P_f)^2}{2\sigma_s^2}} / \sqrt{2\pi(P_f + \sigma_s^2)}} \, dx$$

$$= \frac{1}{2} \left[ \ln \left( 1 + \frac{P_f}{\sigma_s^2} \right) - \frac{P_f}{P_f + \sigma_s^2} \right]$$

While the expression for $D(P_w \| P_s)$ has a closed form, the Taylor series expansion of $D(P_w \| P_s)$ with respect to $P_f$ around $P_f = 0$ is more useful. While the zeroth and first order terms are zero, the second order term is:

$$\frac{P_f^2}{2!} \times \frac{\partial^2 D(P_w \| P_s)}{\partial P_f^2} \bigg|_{P_f=0} = \frac{P_f^2}{4\sigma_w^4}$$

For the third order term we obtain:

$$\frac{P_f^3}{3!} \times \frac{\partial^3 D(P_w \| P_s)}{\partial P_f^3} \bigg|_{P_f=0} = -\frac{P_f^3}{3\sigma_w^6} \leq 0$$
If $P_f < \sigma_w^2$, then the Taylor series converges and we can apply Taylor’s Theorem to upper-bound relative entropy with the second order term. The upper bound we seek is:

$$TV(P_n^w, P_n^s) \leq \frac{P_f}{2\sigma_w^2} \sqrt{\frac{n}{2}}$$ (3)

Suppose Alice sets her average covert symbol power $P_f \leq cf(n)$, where $c = 2\epsilon \sqrt{2}$. In most practical scenarios Alice can lower-bound $\sigma_w^2 \geq \hat{\sigma}_w^2$ and set $f(n) = \hat{\sigma}_w^2$ (a conservative lower bound is the thermal noise power of the best receiver currently available). If $\sigma_w^2$ is unknown, select $f(n)$ such that $f(n) = o(1)$ and $f(n) = \omega(1/\sqrt{n})$ (the latter condition is used to bound Bob’s decoding error probability). In either case, for $n$ large enough, $P_f < \sigma_w^2$ satisfies the Taylor series convergence criterion, and Alice obtains the upper bound $TV(P_n^w, P_n^s) \leq \epsilon$, limiting the performance of Willie’s detector.

Since Alice’s symbol power $P_f$ is a decreasing function of the codeword length $n$, the standard channel coding results for constant power do not directly apply. Thus, we examine the probability $P_e$ of Bob’s decoding error averaged over all possible codebooks. Let Bob employ a maximum-likelihood (ML) decoder (i.e. minimum distance) to process the received vector $y_b$ when $c(W_k)$ was sent. The decoder makes an error when $y_b$ is closer to another codeword $c(W_i)$, $i \neq k$:

$$P_e = \mathbb{E}_{c(W_k)} \left[ P \left( \bigcup_{i=0, i \neq k}^{2^n R} \{ y_b \text{ closer to } c(W_i) \} | c(W_k) \text{ sent} \right) \right]$$

$$\leq \mathbb{E}_{c(W_k)} \left[ \sum_{i=0, i \neq k}^{2^n R} P( y_b \text{ closer to } c(W_i) | c(W_k) \text{ sent} ) \right]$$ (4)

$$= \sum_{i=0, i \neq k}^{2^n R} \mathbb{E}_{c(W_k)} \left[ P( y_b \text{ closer to } c(W_i) | c(W_k) \text{ sent} ) \right]$$ (5)

where (4) follows from the union bound. Let $d = \|c(W_k) - c(W_i)\|_2$ be the distance between two codewords, where $\| \cdot \|_2$ is the $L^2$ norm. Since codewords are independent and Gaussian, $c(W_k) - c(W_i) \sim \mathcal{N}(0, 2P_f)$ and $d^2 = 2P_f U$, where $U \sim \chi_n^2$, with $\chi_n^2$ denoting the chi-squared distribution with $n$ degrees of freedom. Therefore, by [8, (3.44)]:

$$\mathbb{E}_{c(W_k)} \left[ P( y_b \text{ closer to } c(W_i) | c(W_k) \text{ sent} ) \right] = \mathbb{E}_U \left[ Q \left( \sqrt{\frac{P_f U}{2\sigma_b^2}} \right) \right]$$
where \( Q(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt \). Since \( Q(x) \leq \frac{1}{2} e^{-x^2/2} \) [9, (5)] and \( P_f = \frac{c_f(n)}{\sqrt{n}} \cdot \frac{1}{n} \int_{-\infty}^{\infty} e^{-t^2/2} dt \),

\[
E_U \left[ Q\left( \sqrt{\frac{P_f U}{2\sigma_b^2}} \right) \right] \leq E_U \left[ \exp \left( -\frac{c_f(n)U}{4\sqrt{n\sigma_b^2}} \right) \right] \\
= \int_{0}^{\infty} \exp \left( -\frac{c_f(n)u}{4\sqrt{n\sigma_b^2}} \right) \left[ \frac{2^{-n/2}}{\Gamma(n/2)} u^{n/2-1} e^{-u/2} \right] du \\
= 2^{-n/2} \left( \frac{1}{2} + \frac{c_f(n)}{4\sqrt{n\sigma_b^2}} \right)^{-n/2}
\]

(6)

where (6) is from the substitution \( v = u \left( \frac{1}{2} + \frac{c_f(n)}{4\sqrt{n\sigma_b^2}} \right) \) and the definition of the Gamma function \( \Gamma(n) = \int_{0}^{\infty} x^{n-1} e^{-x} dx \). Since \( \frac{1}{2} + \frac{c_f(n)}{4\sqrt{n\sigma_b^2}} = 2^{\log_2 \left( \frac{1}{2} + \frac{c_f(n)}{4\sqrt{n\sigma_b^2}} \right)} \):

\[
E_{c(W_k)} [P(y_b \text{ closer to } c(W_i)|c(W_k) \text{ sent})] \leq 2^{-n/2} \log_2 \left( \frac{1}{2} + \frac{c_f(n)}{4\sqrt{n\sigma_b^2}} \right)
\]

Hence, the summand in (5) does not depend on \( i \), and (5) becomes:

\[
P_e \leq 2^{-nR - \frac{n}{2} \log_2 \left( \frac{1}{2} + \frac{c_f(n)}{4\sqrt{n\sigma_b^2}} \right)}
\]

Since \( f(n) = \omega(1/\sqrt{n}) \), if rate \( R = \frac{\rho}{2} \log_2 \left( 1 + \frac{c_f(n)}{2\sqrt{n\sigma_b^2}} \right) \) for a constant \( \rho < 1 \), as \( n \) increases, the probability of Bob’s decoding error decays exponentially to zero and Bob obtains \( nR = \sqrt{n} \frac{\rho}{2} \log_2 \left( 1 + \frac{c_f(n)}{2\sqrt{n\sigma_b^2}} \right)^{\sqrt{n}} \) covert bits in \( n \) channel uses. Since \( nR \leq \frac{\sqrt{n} \rho cf(n)}{4\sigma_b^2 \ln 2} \), approaching equality as \( n \) gets very large, Bob receives \( o(\sqrt{n}) \) bits in \( n \) channel uses, and \( O(\sqrt{n}) \) bits in \( n \) channel uses if \( f(n) = \hat{\sigma}_w^2 \).

Implications of a peak power constraint

Since most practical systems are peak-power constrained, we show that the square root law holds for the binary input Gaussian output channel using a proof similar to that of Theorem 1.1.

**Theorem 1.2** (Achievability under a peak power constraint). Suppose Alice’s transmitter is subject to the peak power constraint \( b \) and Willie’s channel is subject to AWGN with power \( \sigma_w^2 > 0 \). Then Alice can maintain Willie’s sum of the probabilities of detection errors \( \alpha + \beta \geq 1 - \epsilon \) for any \( \epsilon > 0 \) while covertly transmitting \( o(\sqrt{n}) \) bits to Bob over \( n \) uses of an AWGN channel if \( \sigma_w^2 \) is unknown and \( O(\sqrt{n}) \) bits in \( n \) channel uses if she can lower-bound \( \sigma_w^2 \geq \hat{\sigma}_w^2 \).
**Proof:** **Construction:** Alice encodes the input in blocks of size $M$ bits into codewords of length $n$ at the rate $R = M/n$ bits/symbol with the symbols drawn from alphabet $\{-a, a\}$, where $a$ satisfies the peak power constraint $a^2 < b$ and is defined later. We independently generate $2^{nR}$ codewords $\{c(W_k), k = 1, 2, \ldots, 2^{nR}\}$ for messages $W_k$ from $\{-a, a\}^n$ according to $p_X(x) = \prod_{i=1}^{n} p_X(x_i)$, where $p_X(-a) = p_X(a) = \frac{1}{2}$. As in the proof of Theorem 1.1, the codebook is a secret key shared between Alice and Bob, but Willie knows how it is constructed, including the value of $a$.

**Analysis:** When Alice transmits a covert symbol during the $i^{th}$ symbol period, she transmits $-a$ or $a$ equiprobably by construction and Willie observes the covert symbol corrupted by AWGN. Therefore, $P_s = \frac{1}{2} (\mathcal{N}(-a, \sigma_w^2) + \mathcal{N}(a, \sigma_w^2))$, and, with $P_w = \mathcal{N}(0, \sigma_w^2)$, we have:

$$D(P_w \| P_s) = \int_{-\infty}^{\infty} \frac{e^{-\frac{x^2}{2\sigma_w^2}}}{\sqrt{2\pi\sigma_w}} \ln \left( \frac{e^{-\frac{(x+a)^2}{2\sigma_w^2}} + e^{-\frac{(x-a)^2}{2\sigma_w^2}}}{2} \right) dx$$

There is no closed-form expression for $D(P_w \| P_s)$, but it can be expanded using the Taylor series with respect to $a$ around $a = 0$. While the zeroth through third order terms are zero, the fourth order term is:

$$\frac{a^4}{4!} \times \left. \frac{\partial^4 D(P_w \| P_s)}{\partial a^4} \right|_{a=0} = \frac{a^4}{4\sigma_w^4}$$

While the fifth order term is zero, for the sixth order term we obtain:

$$\frac{a^6}{6!} \times \left. \frac{\partial^6 D(P_w \| P_s)}{\partial a^6} \right|_{a=0} = -\frac{a^6}{3\sigma_w^6} < 0$$

If $a < \sigma_w$, then the Taylor series converges and we can apply Taylor’s Theorem and upper-bound relative entropy with the fourth order term. The upper bound we seek is:

$$TV(P_w^n, P_s^n) \leq \frac{a^2}{2\sigma_w^2} \sqrt{n} \frac{\sqrt{n}}{2}$$

Since the power of Alice’s covert symbol is $a^2 = P_f$, (7) is identical to (3) and Alice sets $a^2 \leq cf(n) / \sqrt{n}$, where $c$ and $f(n)$ are defined as in Theorem 1.1. Then, for $n$ large enough, $a < \sigma_w$ satisfies
the Taylor series convergence criterion, and Alice obtains the upper bound \( TV(P_n^{w}, P_n^{s}) \leq \epsilon \), limiting the performance of Willie’s detector.

Like in Theorem 1.1, we cannot directly apply the standard constant-power channel coding results to our system where the symbol power is a decreasing function of the codeword length. We upper-bound Bob’s decoding error probability by analyzing a suboptimal decoding scheme. Suppose Bob uses a hard-decision device on each received covert symbol \( y_i^{(b)} = f_i + z_i^{(b)} \) via the rule \( \hat{f}_i = \begin{cases} a & \text{if } y_i^{(b)} \geq 0; \\ -a & \text{otherwise} \end{cases} \), and applies an ML decoder on its output. The effective channel for the encoder/decoder pair is a binary symmetric channel with cross-over probability \( p_e = Q(a/\sigma_b) \) and the probability of the decoding error averaged over all possible codebooks is \( P_e \leq 2^{nR-n(1-H(p_e))} \) [10], where \( H(p) = -p \log_2 p - (1-p) \log_2 (1-p) \) is the binary entropy function. We expand the analysis in [11, Section I.2.1] to characterize the rate \( R \). The Taylor series of \( e^{-t^2/2} \) alternates, and the Taylor series expansion of \( p_e = Q(a/\sigma_b) = 1/2 \int_0^a \frac{2}{\sqrt{2\pi}} e^{-t^2/2} dt \) with respect to \( a \) around \( a = 0 \) (which converges since \( a \) is small for large \( n \)) yields an upper bound:

\[
p_e \leq \frac{1}{2} - \frac{1}{\sqrt{2\pi}} \left( \frac{a}{\sigma_b} - \frac{a^3}{6\sigma_b^3} \right) = p_e^{(UB)}.
\]

Since \( H(p) \) is a monotonically increasing function on the interval \([0, \frac{1}{2}]\), \( H(p_e) \leq H(p_e^{(UB)}) \). The odd terms of the Taylor series expansion of \( H(p_e^{(UB)}) \) with respect to \( a \) around \( a = 0 \) are zero, and, thus, \( H(p_e^{(UB)}) = 1 - \frac{a^2}{\sigma_b^2 \pi \ln 2} + \mathcal{O}(a^4) \). Since \( a^2 = \frac{cf(n)}{\sqrt{n}} \), \( P_e \leq 2^{nR-\frac{\sqrt{n}cf(n)}{\sigma_b^2 \pi \ln 2} + \mathcal{O}(1)} \). Since \( f(n) = \omega(1/\sqrt{n}) \), if rate \( R = \frac{\rho cf(n)}{\sqrt{n} \sigma_b^2 \pi \ln 2} \) bits/symbol for a constant \( \rho < 1 \), the probability of Bob’s decoding error decays exponentially to zero as \( n \) increases and Bob obtains \( nR = o(\sqrt{n}) \) bits in \( n \) channel uses, and \( \mathcal{O}(\sqrt{n}) \) bits in \( n \) channel uses if \( f(n) = \delta_{w}^2 \).

Remarks

**Relationship with the Gaussian wire-tap channel [12]:** Consider \( \sigma_b^2 > \sigma_w^2 \). From Theorems 1.1 and 1.2, Alice can send a positive number of bits covertly to Bob; however, the secrecy capacity of the Gaussian wire-tap channel [12] is zero. This seems paradoxical until we consider that in the wire-tap scenario, Alice’s objective is to prevent Willie from decoding her message.
to Bob. She fails when $\sigma_b^2 > \sigma_w^2$ because Willie can decode any message she sends to Bob using public codebooks, as the capacity of Willie’s channel to her is greater than Bob’s. However, here Alice and Bob’s codebook is private and Willie’s ability to distinguish Alice’s transmission from random noise is limited by the sum of the probabilities of his detection errors, which is controlled by Alice employing a constrained transmission power. Thus, provided they agree on a codebook beforehand, Alice can covertly communicate to Bob even when the channel from Alice to Willie is less noisy than the channel from Alice to Bob.

**Relationship with Square Root Law in Steganography:** It has recently been shown that in finite-alphabet imperfect steganographic systems at most $O(\sqrt{n})$ symbols in the original covertext of size $n$ may safely be modified to hide a steganographic message [2]. From the steganographic perspective, our covertext is the noise on Willie’s channel to Alice. However, our result does not obey their converse, as we can modify all symbols in our covertext, highlighting the different nature of the problem scenarios. Nevertheless, it is worthwhile to consider a scenario where roughly $\tau n$ out of $n$ of symbols are used to carry the message.

Let’s construct the codebook in two stages. First, flip a biased coin $n$ times and set the $i^{th}$ symbol in every codeword to one with probability $\tau$ (and zero otherwise). Denote the number of symbols set to one as $\eta$ and note that $\mathbb{E}[\eta] = \tau n$. We complete the codebook by independently generating $2^{nR}$ vectors of length $\eta$ according to $p_X(x) = \prod_{i=1}^{\eta} p_X(x_i)$, where $X \sim \mathcal{N}(0, P_f)$, and assigning the values of these vectors to the cells in the corresponding codewords that contain ones. Thus, a codeword uses $\tau n$ symbols on average (over all codebooks), and all codewords contain zeros in identical locations, facilitating decoding for Bob. The coin flip is independent of both the symbol and the channel noise. When Alice is transmitting a codeword, the distribution of each of Willie’s observations is $P_s = (1-\tau)\mathcal{N}(0, \sigma_w^2) + \tau \mathcal{N}(0, P_f + \sigma_w^2)$ and, thus,

$$D(P_w \| P_s) = \int_{-\infty}^{\infty} \frac{e^{-\frac{x^2}{2\sigma_w^2}}}{\sqrt{2\pi \sigma_w^2}} \ln \frac{e^{-\frac{x^2}{2\sigma_w^2}}/\sqrt{2\pi \sigma_w^2}}{(1-\tau)e^{-\frac{x^2}{2\sigma_w^2}}/\sqrt{2\pi \sigma_w^2} + \tau e^{-\frac{2(P_f + \sigma_w^2)}{\sqrt{2\pi (P_f + \sigma_w^2)}}}} \, dx$$

There is no closed-form expression for $D(P_w \| P_s)$, but a Taylor series expansion with respect
to $P_f$ around $P_f = 0$ yields the following bound:

$$TV(P^n_w, P^n_s) \leq \frac{\tau P_f}{2\sigma^2_w} \sqrt{\frac{n}{2}} \quad (8)$$

The only difference in (8) from (3) is $\tau$ in the numerator. Thus, if Alice sets the product $\tau P_f = \frac{c f(n)}{\sqrt{n}}$, with $c$ and $f(n)$ as previously defined, she limits the performance of Willie’s detector. This product is the average symbol power used by Alice. It is easy to verify that in the peak power constrained scenario Alice should set product $\tau a^2 = \frac{c f(n)}{\sqrt{n}}$ and that the number of bits that Alice can covertly transmit to Bob obeys the previously derived square root bounds. This demonstrates the richness of our scenario and the generality of our square root law.

IV. CONVERSE

Here, as in the achievability, the channel between Alice and Bob is subject to AWGN of power $\sigma^2_b$. Alice’s objective is to covertly transmit a message $W_k$ that is $M = \omega(\sqrt{n})$ bits long to Bob in $n$ channel uses with arbitrarily small probability of decoding error as $n$ gets large. Alice encodes each message $W_k$ arbitrarily into $n$ symbols at the rate $R = M/n$ symbols/bit. For an upper bound on the reduction in entropy, the messages are chosen equiprobably.

Willie observes all $n$ of Alice’s channel uses. To strengthen the converse, he is oblivious to her signal properties. Nevertheless, even with Willie’s knowledge limited, Alice cannot transmit a message with $\omega(\sqrt{n})$ bits of information in $n$ channel uses without either being detected by Willie or having Bob suffer a non-zero decoding error.

**Theorem 2.** If over $n$ channel uses, Alice attempts to transmit a covert message to Bob that is $\omega(\sqrt{n})$ bits long, then, as $n \to \infty$, either Willie can detect her with arbitrarily low sum of error probabilities $\alpha + \beta$, or Bob cannot decode with arbitrarily low probability of error.

**Proof:** To detect Alice’s covert transmissions, Willie performs the following hypothesis test:

$$H_0 : \quad y_i^{(w)} = z_i^{(w)}, \quad i = 1, \ldots, n$$
$$H_1 : \quad y_i^{(w)} = f_i + z_i^{(w)}, \quad i = 1, \ldots, n$$

Rejection of $H_0$ means that Alice is covertly communicating with Bob. First, we show how Willie can bound the errors $\alpha$ and $\beta$ of this test as a function of Alice’s signal parameters. Then we show that if Alice prevents Willie’s test from detecting her by adjusting her signal power, Bob cannot decode her transmissions without error.

To perform the test, Willie collects a vector of $n$ independent readings $\mathbf{y}_w$ from his channel to Alice and generates the test statistic $S = \frac{\mathbf{y}_w \mathbf{x}_w^T}{n}$ where $\mathbf{x}_w^T$ denotes transpose of vector $\mathbf{x}$. Under the null hypothesis $H_0$ Alice does not transmit and Willie reads AWGN on his channel. Thus, $y_i^{(w)} \sim \mathcal{N}(0, \sigma_w^2)$, and the mean and the variance of $S$ when $H_0$ is true are:

$$\mathbf{E}[S] = \sigma_w^2$$  
(9)

$$\text{var}[S] = \frac{2\sigma_w^4}{n}$$  
(10)

Suppose Alice transmits codeword $c(W_k) = \{f_i^{(k)}\}_{i=1}^n$. Then Willie’s vector of observations $\mathbf{y}_{w,k} = \{y_i^{(w,k)}\}_{i=1}^n$ contains readings of mean-shifted noise $y_i^{(w,k)} \sim \mathcal{N}(f_i^{(k)}, \sigma_w^2)$. The mean of each squared observation is $\mathbf{E}[y_i^2] = \sigma_w^2 + (f_i^{(k)})^2$, while the variance is $\text{var}[y_i^2] = \mathbf{E}[y_i^4] - (\mathbf{E}[y_i^2])^2 = 4(f_i^{(k)})^2 \sigma_w^2 + 2\sigma_w^4$. Denote the average power per symbol of codeword $c(W_k)$ by $P_k = \frac{c(W_k)c^T(W_k)}{n}$. Then the mean and variance of $S$ when Alice transmits message $W_k$ are:

$$\mathbf{E}[S] = \sigma_w^2 + P_k$$  
(11)

$$\text{var}[S] = \frac{4P_k\sigma_w^2 + 2\sigma_w^4}{n}$$  
(12)

The variance of Willie’s test statistic (12) is computed by adding the variances conditioned on $c(W_k)$ of the squared individual observations $\text{var}[y_i^2]$ (and dividing by $n^2$) since the noise on the individual observations is independent.

The probability distribution for the vector of Willie’s observations depends on which hypothesis is true. Denote $\mathbf{P}_0$ as the distribution when $H_0$ holds, and $\mathbf{P}_1^{(k)}$ when $H_1$ holds with Alice transmitting message $W_k$. While $\mathbf{P}_1^{(k)}$ is conditioned on Alice’s codeword, we show that the power of the codeword determines its detectability by this detector, and that our result applies to all codewords with power of the same order.
If $H_0$ is true, then $S$ should be close to (9). Willie picks some threshold $t$ and compares the value of $S$ to $\sigma^2_w + t$. He accepts $H_0$ if $S < \sigma^2_w + t$ and rejects it otherwise. Suppose that he desires false positive probability $\alpha^*$, which is the probability that $S \geq \sigma^2_w + t$ when $H_0$ is true. We bound it using (9) and (10) with Chebyshev’s Inequality [7, (3.32)]:

$$\alpha = P_0(S \geq \sigma^2_w + t) \leq P_0(|S - \sigma^2_w| \geq t) \leq \frac{2\sigma^4_w}{nt^2}$$

Thus, to obtain $\alpha^*$, Willie sets $t = \frac{d}{\sqrt{n}}$, where $d = \sqrt{\frac{2\sigma^2_w}{\sigma^*}}$ is a constant. As $n$ increases, $t$ decreases, which is consistent with Willie gaining greater confidence with more observations.

Suppose Alice transmits $W_k$. Then the probability of a miss $\beta^{(k)}$ given $t$ is the probability that $S < \sigma^2_w + t$, which we bound using (11) and (12) with Chebyshev’s Inequality:

$$\beta^{(k)} = P_1^{(k)}(S < \sigma^2_w + t) \leq P_1^{(k)}(|S - \sigma^2_w - P_k| \geq P_k - t) \leq \frac{4P_k\sigma^2_w + 2\sigma^4_w}{(\sqrt{n}P_k - d)^2}$$

If $P_k = \omega(1/\sqrt{n})$, $\lim_{n \to \infty} \beta^{(k)} = 0$. Thus, with enough observations, Willie can detect with arbitrarily low error probability Alice’s codewords with average symbol power $P_k = \omega(1/\sqrt{n})$. Note that Willie’s detector is oblivious to any details of Alice’s codebook construction.

By (13), if Alice desires to lower-bound the sum of the probabilities of error of Willie’s statistical test by $\alpha + \beta \geq \zeta > 0$, she must use low-power codewords; in particular, a fraction $\gamma > 0$ of the codewords must have $P_{\text{td}} = O(1/\sqrt{n})$. Let’s denote this set of codewords as $\mathcal{U}$ and examine the probability of Bob’s decoding error $P_e$. The probability that a message from set $\mathcal{U}$ is sent is $P(\mathcal{U}) = \gamma$, as all messages are equiprobable. We bound $P_e = P_e(\mathcal{U})P(\mathcal{U}) + P_e(\overline{\mathcal{U}})P(\overline{\mathcal{U}}) \geq \gamma P_e(\mathcal{U})$, where $\overline{\mathcal{U}}$ is the complement of $\mathcal{U}$ and $P_e(\mathcal{U})$ is the probability of decoding error when a message from $\mathcal{U}$ is sent:

$$P_e(\mathcal{U}) = \frac{1}{|\mathcal{U}|} \sum_{W \in \mathcal{U}} P_e(c(W) \text{ transmitted})$$

where $P_e(c(W) \text{ transmitted})$ is the probability of error when codeword $c(W)$ is transmitted, $|\cdot|$ denotes the set cardinality operator, and (14) holds because all messages are equiprobable.

When Bob uses the optimal decoder, $P_e(c(W) \text{ transmitted})$ is the probability that Bob decodes the received signal as $\hat{W} \neq W$. This is the probability of a union of events $E_j$, where $E_j$ is the
event that sent message $W$ is decoded as some other message $W_j \neq W$:

$$P_e(e(W) \text{ transmitted}) = P \left( \bigcup_{j=1}^{\gamma 2^n} W_j \neq W \big| E_j \right) \geq P \left( \bigcup_{W_j \in \mathcal{U} \setminus \{W\}} W_j \right) \triangleq P_e^{(U)}$$

where the inequality in (15) is since the sets in the second union are contained in the first. From the decoder perspective, this is due to the decrease in the decoding error probability is Bob knew that the message came from $\mathcal{U}$ (reducing the set of messages on which the decoder can err).

Our analysis of $P_e^{(U)}$ uses Cover’s simplification of Fano’s inequality similar to the proof of the converse to the coding theorem for Gaussian channels in [7, Ch. 9.2]. Since we are interested in $P_e^{(U)}$, we do not absorb it into $\epsilon$ as done in (9.37) of [7]. Rather, we explicitly use:

$$H(W|\hat{W}) \leq 1 + \left( \log_2 |\mathcal{U}| \right) P_e^{(U)}$$

where $H(W|\hat{W})$ denotes the entropy of message $W$ conditioned on Bob’s decoding $\hat{W}$ of $W$.

Noting that the size of the set $\mathcal{U}$ from which the messages are drawn is $\gamma 2^n R$ and that, since each message is equiprobable, the entropy of a message $W$ from $\mathcal{U}$ is $H(W) = \log_2 |\mathcal{U}| = \log_2 \gamma + nR$, we utilize (16) and carry out steps (9.38)–(9.53) in [7] to obtain:

$$P_e^{(U)} \geq 1 - \frac{P_{\mathcal{U}}/2\sigma_b^2 + 1/n}{\log_2 \gamma + R}$$

Since Alice transmits $\omega(\sqrt{n})$ bits in $n$ channel uses, her rate is $R = \omega(1/\sqrt{n})$ bits/symbol. However, $P_\mathcal{U} = O(1/\sqrt{n})$, and, as $n \to \infty$, $P_e^{(U)}$ is bounded away from zero. Since $\gamma > 0$, $P_e$ is bounded away from zero if Alice tries to beat Willie’s simple hypothesis test.

**Goodput of Alice’s Communication**

Define the goodput $G(n)$ of Alice’s communication as the average number of bits that Bob can receive from Alice over $n$ channel uses with non-zero probability of a message being undetected as $n \to \infty$. Since only $\mathcal{U}$ contains such messages, by (17), the probability of her message being successfully decoded by Bob is $P_s^{(U)} = 1 - P_e^{(U)} = O \left( \frac{1}{\sqrt{n}R} \right)$ and the goodput is $G(n) = \gamma P_s^{(U)} Rn = O(\sqrt{n})$. Thus, Alice cannot break the square root law using an arbitrarily high transmission rate while keeping the power under Willie’s detection threshold.
V. DISCUSSION

A. Mapping to Continuous-time Channel

We employ a discrete-time model throughout the paper. However, whereas this is a common assumption made without loss of generality in standard communication theory, it is important to consider whether some aspect of the LPD problem has been missed by starting in discrete-time.

Consider the standard communication system model, where Alice’s (baseband) continuous-time waveform would be given in terms of her discrete-time transmitted sequence by:

\[ x(t) = \sum_{i=1}^{n} f_i \ p(t - iT_s) \]

where \( T_s \) is the symbol period and \( p(\cdot) \) is the pulse shaping waveform. Consider a (baseband) system bandwidth constraint of \( W \) Hz. Now, if Alice chooses \( p(\cdot) \) ideally as sinc\((2Wt)\), where sinc\((x) = \frac{\sin(\pi x)}{\pi x}\), then the natural choice of \( T_s = 1/2W \) results in no intersymbol interference (ISI). From the Nyquist sampling criterion, both Willie (and Bob) can extract all of the information from the signaling band by sampling at a rate of \( 2W \) samples/second, which then leads directly to the discrete-time model of Section II and suits our demonstration of the fundamental limits to Alice’s covert channel capabilities. However, when \( p(\cdot) \) is chosen in a more practical fashion, for example, as a raised cosine pulse with some excess bandwidth, then sampling at a rate higher than \( 2W \) has utility for signal detection even if the Nyquist ISI criterion is satisfied. In particular, techniques involving cyclostationary detection are now applicable, and we consider such a scenario a promising area for future work.

B. Fading and Shadowing

Fading and shadowing will impact both the capacity of the channel from Alice to Bob and the ability for Willie to detect Alice’s transmission. There are a number of different models that could be employed to incorporate these effects. However, while these models will have a significant impact as we move toward practical systems, they are unlikely to have an impact on the asymptotic results presented here.
C. Relationship to Previous Work

The LPD communication problem is related to the problem of establishing a cognitive radio (CR) network [13]. An aspect of the CR problem is limiting the interference from the secondary users’ radios to the primary users of the network. The LPD problem with a passive warden can be cast within this framework by having primary users only listen [14]. However, the properties of the secondary signal that allows smooth operation of the primary network are very different from those of an undetectable signal. While there is a lot of work on the former topic, we are not aware of work by the CR community on the latter issue.

Analytical evaluation of LPD communication has been sparse. Hero studies LPI/LPD channels [1] in a multiple-input multiple-output (MIMO) setting. However, he focuses on the constraints (s.t. power, fourth moment, etc.) that the LPD communication over a MIMO channel should enforce given the kind of information the adversary possesses and on the signaling methods that maximize the throughput of the channel given those constraints. While he recognizes that an LPD communication system is constrained by average power, he does not analyze the constraint asymptotically and, thus, does not obtain the square root law. It is notable that the LPI portion of his work has drawn significant attention, while the LPD portion has been largely overlooked.

Unlike LPD communication, much analytical work has been done on steganography. An excellent survey of work prior to 1999 is provided by Petitcolas [15]. A lot of the research effort in this area focuses on measuring the security of steganographic systems, which is particularly important for imperfect steganography as it allows the user to quantify the risk of being detected. Proposals for measures of security include relative entropy [16], Fisher information [17], as well as the metric of [18] which is similar to the sum of Willie’s detection errors that we employ. As noted in the remark in Section III, the square root law was found in finite-alphabet imperfect steganography [2]. However, although their goal is the same as ours (hiding information with low probability of detection by Willie), their model based on hiding information in finite-alphabet images is very different from ours. As demonstrated by the constructions of Section III, our
scenario is arguably richer, and its additional degree of freedom in the choice of transmission power allows Alice to alter all $n$ symbols used in transmission while maintaining a fixed detection probability, which stands in contrast to the finite-alphabet steganography result.

VI. CONCLUSION

Practitioners have always known that LPD communication requires one to use low power in order to blend in with the noise on the eavesdropping warden’s channel. However, the specific requirements for achieving LPD communication and resulting achievable performance have seldom been analyzed prior to this work. We quantified the conditions for existence and maintenance of an LPD channel by proving that the LPD communication is subject to a square root law in that the number of bits that can be covertly transmitted in $n$ channel uses is $O(\sqrt{n})$. An interesting result in our work is the fact that one can use all of the $n$ symbols with positive power to transmit the covert messages.

There are a number of avenues for future research. Practical network settings and the implications of the square root law on the covert transmission of packets under additional constraints such as delay should be analyzed. The impact of dynamism in the network should also be examined, as well as more realistic scenarios that include channel artifacts such as fading and interference from other nodes. One may be able to improve LPD communication by employing nodes that perform friendly jamming. Eventually, we would like to answer this fundamental question: is it possible to establish and maintain a “shadow” wireless network in the presence of both active and passive wardens?

REFERENCES


