LPD Communication when the Warden Does Not Know When

Boulat A. Bash*, Dennis Goeckel†, Don Towsley*

*School of Computer Science, University of Massachusetts, Amherst, Massachusetts 01003–9264
†Electrical and Computer Engineering Department, University of Massachusetts, Amherst, Massachusetts 01003–9292

***Abstract*—Unlike standard security methods (e.g. encryption), *low probability of detection* (LPD) communication does not merely protect the information contained in a transmission from unauthorized access, but prevents the detection of a transmission in the first place. In this work we study the impact of secretly pre-arranging the time of communication. We prove that if Alice has AWGN channels to Bob and the warden, and if she and Bob can choose a single n symbol period slot out of $T(n)$ such slots, keeping the selection secret from the warden (and, thus, forcing him to monitor all $T(n)$ slots), then Alice can reliably transmit $\mathcal{O}(\min\{\sqrt{n \log T(n)}, n\})$ bits to Bob while keeping the warden's detector ineffective. The result indicates that only an additional $\log T(n)$ secret bits need to be exchanged between Alice and Bob prior to communication to produce a multiplicative gain of $\sqrt{\log T(n)}$ in the amount of transmitted covert information.**

I. Introduction

Edward Snowden's recent revelations of massive surveillance programs by the US National Security Agency (NSA) have emphasized the necessity of secure communication systems that do not just protect the content of the user's message from being decoded, but prevent the detection of its transmission in the first place. Since encrypted data or even just the transmission of a signal can arouse suspicion, and even the most theoretically robust cryptographic security scheme can be defeated by a determined adversary using non-computational methods such as side-channel analysis, *low probability of detection* (LPD) communication has substantial application.

In the LPD communication scenario, Alice transmits a message to Bob over a noisy channel while the warden Willie attempts to detect her transmission. The channel from Alice to Willie is also subject to noise. Thus, while Alice transmits low-power covert signals to Bob, Willie attempts to classify these signals as either noise on his channel or signals from Alice. Recent work on LPD communication [1]–[3] characterizes the amount of information that can be transmitted by Alice to Bob with Alice tolerating a small probability of being detected by Willie. However, these studies assume that Willie knows when Alice starts transmitting (if she transmits).

There are many practical scenarios where Alice and Bob have a pre-arranged time of communication that is unknown to Willie (e.g. a certain time on a given day). Alice's message may also be much shorter than the total time available to transmit it (e.g. a few seconds out of the day when both Alice and Bob are available). Thus, since Willie does not know when Alice transmits, he has to monitor a much longer time period than the duration of Alice's transmission. In this work we show how Alice can leverage Willie's ignorance of her transmission time to transmit significant additional information to Bob.

This research was sponsored by the National Science Foundation under grants CNS-1018464, CNS-0964094, and ECCS-1309573.

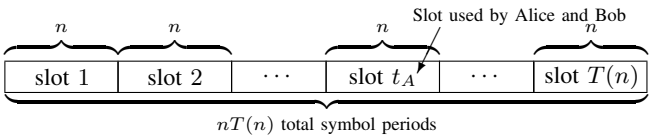

Fig. 1. Slotted channel: each of the $T(n)$ slots contains n symbol periods. Alice and Bob use slot t_A to communicate.

In our scenario, Alice communicates with Bob over an additive white Gaussian noise (AWGN) channel. Willie also has an AWGN channel from Alice. Unlike the setting in [1], [2], the channel is slotted, as described in Figure 1. Each of $T(n)$ slots contains n symbol periods, where $T(n)$ is an increasing function of n. If Alice used all $nT(n)$ symbol periods for transmission, then, by [1], [2], she could reliably transmit $\mathcal{O}(\sqrt{nT(n)})$ LPD bits to Bob. However, Alice uses only a single slot t_A. She agrees on the slot with Bob but keeps it secret from Willie, who is thus forced to monitor all $T(n)$ slots. A naïve application of the square root law from [1], [2] allows Alice to reliably transmit $\mathcal{O}(\sqrt{n})$ covert bits in this scenario. We demonstrate that Alice can transmit $\mathcal{O}\left(\min\{\sqrt{n \log T(n)}, n\}\right)$ bits reliably on this channel while maintaining arbitrarily low probability of detection by Willie. Conversely, we show that the transmission of $\omega(\sqrt{n \log T(n)})$ bits either results in Alice being detected with high probability or unreliable communication.

The cost of LPD communication on the AWGN channel is the secret that Alice and Bob share before the transmission. Remarkably, our results here demonstrate that the additive expense of an extra $\log T(n)$ secret bits to indicate the slot employed by Alice and Bob esults in a *multiplicative* increase (by a factor of $\sqrt{\log T(n)}$) in the number of covert bits that Alice can transmit reliably to Bob. Timing is thus a very efficient resource for LPD communication. It also necessitates vastly different analysis than the power-based LPD communication examined in [1], [2]. Specifically, the relative entropy based bounds on the probability of detection error employed in [1], [2] are too loose to yield our achievability

results, and we therefore have to apply other techniques from mathematical statistics.

We state our main result using asymptotic notation where $f(n) = \mathcal{O}(g(n))$ and $f(n) = \Omega(g(n))$ denote asymptotically tight upper and lower bounds on $f(n)$, respectively, and $f(n) = o(g(n))$ and $f(n) = \omega(g(n))$ denote upper and lower bounds, respectively, that are not asymptotically tight [4, Ch. 3.1]:

Theorem. *Suppose the channel between Alice and each of Bob and Willie experiences independent additive white Gaussian noise (AWGN) with constant power $\sigma_b^2 > 0$ and $\sigma_w^2 > 0$, respectively. Also suppose that Alice transmits on one of $T(n)$ slots chosen randomly. Each slot contains n symbol periods, where $T(n) = \omega(1)$. Then, for any $\epsilon > 0$, there exists n_0 such that, for all $n \geq n_0$, Alice can reliably transmit $\mathcal{O}\left(\min\{\sqrt{n \log T(n)}, n\}\right)$ bits to Bob in a selected slot while maintaining a probability of detection error by Willie greater than $\frac{1}{2} - \epsilon$. Conversely, if Alice tries to transmit $\omega(\sqrt{n \log T(n)})$ bits on an arbitrary slot, either Willie detects with arbitrarily low probability of error or Bob cannot decode her message with arbitrary low probability of decoding error.*

After introducing our slotted channel model and hypothesis testing background in Section II, we prove the achievability and the converse in Sections III and IV, respectively. We discuss the relationship of this work to steganography in Section V, and conclude in Section VI.

II. Prerequisites

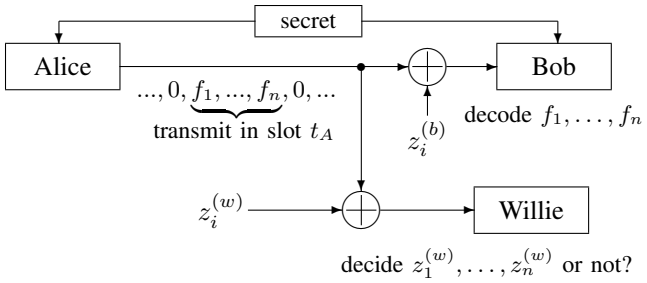

Fig. 2. System framework: Alice and Bob share a secret and agree on a slot t_A before transmission. Alice encodes information into a vector of real symbols $\mathbf{f} = \{f_i\}_{i=1}^n$ and uses slot t_A to transmit it on an AWGN channel to Bob. Upon observing the channel from Alice, Willie has to classify his vector of readings $\mathbf{y}_w$ as either an AWGN vector $\mathbf{z}_w = \{z_i^{(w)}\}_{i=1}^{nT(n)}$ or an AWGN vector that contains a slot with transmissions corrupted by AWGN.

A. Channel Model

We use the discrete-time slotted AWGN channel model with real-valued symbols depicted in Figures 1 and 2. The channel has $T(n)$ slots, each containing n symbol periods. Alice and Bob secretly agree on a slot t_A prior to transmission and Alice uses it to transmit a vector of n real-valued symbols $\mathbf{f} = \{f_i\}_{i=1}^n$. Bob only listens to slot t_A and receives a vector $\mathbf{y}_b = \{y_i^{(b)}\}_{i=1}^n$ 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 $\mathbf{y}_w = \{\mathbf{y}_w(t)\}_{t=1}^{T(n)}$ where $\mathbf{y}_w(t) = [y_{(t-1)n+1}^{(w)}, \ldots, y_{tn}^{(w)}]$ is a vector of observations of slot t, $y_{(t_A-1)n+i}^{(w)} = f_i + z_{(t_A-1)n+i}^{(w)}$ and $y_{(t-1)n+i}^{(w)} = z_{(t-1)n+i}^{(w)}$ for all $t \neq t_A$. Here $\{z_i^{(w)}\}_{i=1}^{nT(n)}$ is an i.i.d. sequence with $z_i^{(w)} \sim \mathcal{N}(0, \sigma_w^2)$. Willie does not know t_A and has to test the entire $\mathbf{y}_w$ to determine whether Alice is communicating. We discuss statistical hypothesis tests next.

B. Hypothesis Testing

Willie expects the observation vector $\mathbf{y}_w$ to be generated by the noise on his channel. He performs a statistical hypothesis test [5] on $\mathbf{y}_w$, where the null hypothesis H_0 is that Alice does not transmit and each sample is i.i.d. $y_i^{(w)} \sim \mathcal{N}(0, \sigma_w^2)$. The alternate hypothesis H_1 is that Alice transmits, and samples from one of the slots come from a different distribution. The rejection of H_0 when it is true is a false alarm (FA) and the acceptance of H_0 when it is false is a missed detection (MD). The probability of error $\mathbb{P}_e^{(w)}$ is the sum of these error probabilities weighted by the prior probabilities of the hypotheses. We assume that Willie is ignorant of the likelihood of Alice transmitting, and that the prior probabilities $\mathbb{P}(H_0 \text{ true}) = \mathbb{P}(H_1 \text{ true}) = \frac{1}{2}$. Thus, $\mathbb{P}_e^{(w)} = \frac{\mathbb{P}_{FA} + \mathbb{P}_{MD}}{2}$.

III. Achievability

Theorem 1 (Achievability)**.** *Suppose Alice has a slotted AWGN channel to Bob with $T(n) = \omega(1)$ slots, each containing n symbol periods. Then, provided that Alice and Bob have a secret of sufficient length, Alice can reliably transmit $\mathcal{O}\left(\min\{\sqrt{n \log T(n)}, n\}\right)$ bits in a single slot while $\lim_{n\to\infty} \mathbb{P}_e^{(w)} > \frac{1}{2} - \epsilon$ for arbitrary $\epsilon > 0$.*

Proof: **Construction:** Alice and Bob secretly select slot t_A uniformly at random out of the $T(n)$ slots in which to communicate. Alice's channel encoder takes as input blocks of length M bits and encodes them into codewords of length n at a rate of $R = M/n$ bits/symbol. We employ random coding arguments and independently generate 2^{nR} codewords $\{\mathbf{c}(W_k), k = 1, 2, \ldots, 2^{nR}\}$ from $\mathbb{R}^n$ for messages $\{W_k\}_{k=1}^{2^{nR}}$, each according to $p_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^n p_X(x_i)$, where $X \sim \mathcal{N}(0, P_f)$ and symbol power $P_f < \frac{\sigma_w^2}{2}$ is defined later. The codebook is used only to send a single message and, along with t_A, is the secret not revealed to Willie, though he knows how it is constructed, including the value P_f.

Another way of viewing the construction is as a choice of one of $T(n)$ codebooks, where the i^{th} codebook has a block of non-zero symbols in the i^{th} slot. Agreement on the timing is equivalent to selection of the t_A^{th} codebook and the message is encoded by choosing a codeword from the selected codebook.

Analysis (Willie): Willie is interested in performing the following hypothesis test on his vector of observations $\mathbf{y}_w$:

H_0 : Alice does not transmit
H_1 : $\exists$ a slot $t_A \in \{1, \ldots, T(n)\}$ in which Alice transmits

Let $Y_t = \sum_{y_i \in \mathbf{y}_w(t)} y_i^2$ be the power in slot t. Since Willie's channel from Alice is corrupted by AWGN with power σ_w^2,

the likelihood function of the observations $\mathbf{y}_w$ under H_0 is:

$$f_0(\mathbf{y}_w) = \left(\frac{1}{2\pi\sigma_w^2}\right)^{\frac{nT(n)}{2}} \exp\left[-\frac{1}{2\sigma_w^2}\sum_{t=1}^{T(n)} Y_t\right]. \quad (1)$$

Since Willie does not know which of the $T(n)$ slots Alice and Bob randomly select for communication, nor the codebook they use, but knows that Alice's signal is Gaussian, the likelihood function of the observations $\mathbf{y}_w$ under H_1 is:

$$f_1(\mathbf{y}_w) = \frac{1}{A(n)(2\pi)^{\frac{nT(n)}{2}}T(n)} \sum_{t=1}^{T(n)} e^{-\frac{Y_t}{2(\sigma_w^2+P_f)}-\frac{B(t)}{2\sigma_w^2}} \quad (2)$$

where $A(n) = \sigma_w^{(T(n)-1)n}(\sigma_w^2+P_f)^{\frac{n}{2}}$ and $B(t) = \sum_{\substack{r=1\\r\neq t}}^{T(n)} Y_r$.

Since the test is between two simple hypotheses on Alice's transmission state, the likelihood ratio test (LRT) is optimal under the Neyman-Pearson criterion [5, Ch. 3.2]. Taking the ratio between (1) and (2), and re-arranging terms, we obtain:

$$\Lambda(\mathbf{y}_w) = \frac{f_1(\mathbf{y}_w)}{f_0(\mathbf{y}_w)} = \left(\frac{\sigma_w^2}{\sigma_w^2+P_f}\right)^{\frac{n}{2}} \frac{1}{T(n)} \sum_{t=1}^{T(n)} e^{\frac{P_f Y_t}{2\sigma_w^2(\sigma_w^2+P_f)}} \quad (3)$$

The likelihood ratio $\Lambda(\mathbf{y}_w)$ is compared to a threshold $\tau(n)$, which is a function of the information known to Willie, and H_0 or H_1 is chosen based on whether $\Lambda(\mathbf{y}_w)$ is smaller or larger than $\tau(n)$ (if it is equal, a random decision is made):

$$\Lambda(\mathbf{y}_w) \underset{H_1}{\overset{H_0}{\lessgtr}} \tau(n) \quad (4)$$

When Alice does not transmit on the i^{th} symbol period, $y_i \sim \mathcal{N}(0,\sigma_w^2)$ since Willie observes AWGN; when Alice transmits, $y_i \sim \mathcal{N}(0,\sigma_w^2+P_f)$ by construction. Let $\{X_t\}$, $X_t \sim \chi_n^2$, $t = 1,\ldots,T(n)$ be a sequence of i.i.d. chi-squared random variables with n degrees of freedom. Then $Y_t = \sigma_w^2 X_t$ for all $t \in \{1,\ldots,T(n)\}$ under H_0 and $t \in \{1,\ldots,T(n)\}\setminus\{t_A\}$ under H_1. However, under H_1, $Y_{t_A} = (\sigma_w^2+P_f)X_{t_A}$.

Consider a random variable $L^{(n)}$ defined as follows:

$$L^{(n)} = \frac{M(n)T(n)\Lambda(\mathbf{y}_w) - (T(n)-1)M(n)}{\sqrt{V(n)}} \quad (5)$$

where $M(n) = \left(\frac{\sigma_w^2+P_f}{\sigma_w^2}\right)^{\frac{n}{2}}$ and

$$V(n) = (T(n)-1)\left[\left(\frac{\sigma_w^2+P_f}{\sigma_w^2-P_f}\right)^{\frac{n}{2}} - \left(\frac{\sigma_w^2+P_f}{\sigma_w^2}\right)^{n}\right]. \quad (6)$$

This is just a deterministically re-normalized LRT statistic. Since n, $T(n)$, σ_w^2, and P_f are known to Willie, and $M(n)$ and $V(n)$ are deterministic functions, the hypothesis test:

$$L^{(n)} = \frac{\sum_{t=1}^{T(n)} e^{\frac{P_f Y_t}{2\sigma_w^2(\sigma_w^2+P_f)}} - (T(n)-1)M(n)}{\sqrt{V(n)}} \underset{H_1}{\overset{H_0}{\lessgtr}} S(n) \quad (7)$$

is equivalent to that in (4), with the threshold

$$S(n) = \frac{M(n)T(n)\tau(n) - (T(n)-1)M(n)}{\sqrt{V(n)}}. \quad (8)$$

The performance of both tests is equal. The probability of error is thus

$$\mathbb{P}_e^{(w)} = \frac{\mathbb{P}(L^{(n)} > S(n)|H_0 \text{ true}) + \mathbb{P}(L^{(n)} \leq S(n)|H_1 \text{ true})}{2} \quad (9)$$

When H_0 is true, we can write (7) as the normalized sum of $T(n)-1$ i.i.d. random variables U_t and an independent random variable $U_{T(n)}/\sqrt{V(n)}$ as follows:

$$L^{(n)} = \frac{1}{\sqrt{V(n)}} \sum_{t=1}^{T(n)-1} (U_t - M(n)) + \frac{U_{T(n)}}{\sqrt{V(n)}}, \quad (10)$$

where $U_{T(n)}$ is identical to U_t that is defined as

$$U_t = \exp\left[\frac{P_f X_t}{2(\sigma_w^2+P_f)}\right]. \quad (11)$$

When H_1 is true, we can write (7) as the normalized sum of $T(n)-1$ i.i.d. random variables U_t and an independent random variable $U_{t_A}/\sqrt{V(n)}$ as follows:

$$L^{(n)} = \frac{1}{\sqrt{V(n)}} \sum_{t=1, t\neq t_A}^{T(n)} (U_t - M(n)) + \frac{U_{t_A}}{\sqrt{V(n)}}, \quad (12)$$

where U_t in the sum is defined as in (11), and

$$U_{t_A} = \exp\left[\frac{P_f X_{t_A}}{2\sigma_w^2}\right]. \quad (13)$$

We first show that the normalized sums in (10) and (12) contain i.i.d. zero-mean unit-variance random variables, thus both converging in distribution to the standard Gaussian distribution $\mathcal{N}(0,1)$ by the central limit theorem (CLT). We then show that, outside the sums, $U_{T(n)}/\sqrt{V(n)} \xrightarrow{\mathcal{P}} 0$ and $U_{t_A}/\sqrt{V(n)} \xrightarrow{\mathcal{P}} 0$, where $K_n \xrightarrow{\mathcal{P}} Q$ denotes convergence of K_n to Q in probability. This allows us to lower bound Willie's probability of error for all values of threshold $S(n)$.

First let's calculate the moments of U_t defined in (11). The expectation of U_t is the moment generating function (MGF) $\mathcal{M}_{\chi_n^2}(x) = (1-2x)^{-n/2}$ of a chi-squared random variable evaluated at $x = \frac{P_f}{2(\sigma_w^2+P_f)}$:

$$\mathbb{E}[U_t] = \mathbb{E}\left[\exp\left(\frac{P_f X_t}{2(\sigma_w^2+P_f)}\right)\right] = \left(\frac{\sigma_w^2+P_f}{\sigma_w^2}\right)^{\frac{n}{2}} \quad (14)$$

Thus, $M(n) = \mathbb{E}[U_t]$, and the terms inside the sum in (10) and (12) have zero mean. The second moment of U_t is:

$$\mathbb{E}[U_t^2] = \mathbb{E}\left[\exp\left(\frac{P_f X_t}{\sigma_w^2+P_f}\right)\right] = \left(\frac{\sigma_w^2+P_f}{\sigma_w^2-P_f}\right)^{\frac{n}{2}} \quad (15)$$

Thus $V(n) = (T(n)-1)\operatorname{Var}[U_t]$, and, by the Lindenberg CLT for a triangular array [6, Th. 27.2], the normalized sums in both (10) and (12) converge in distribution to $\mathcal{N}(0,1)$.

Probability that the magnitude of $\frac{U_{T(n)}}{\sqrt{V(n)}}$ in (10) exceeds $\delta > 0$ is upper-bounded using the Chebyshev's inequality:

$$\mathbb{P}\left(\left|\frac{U_{T(n)}}{\sqrt{V(n)}}\right| > \delta\right) \leq \left(\delta\sqrt{T(n)-1} - R(n)\right)^{-2} \quad (16)$$

where $R(n) = \frac{\mathbb{E}[U_t]}{\sqrt{\mathrm{Var}[U_t]}} = \left[\left(\frac{\sigma_w^4}{\sigma_w^4 - P_f^2}\right)^{\frac{n}{2}} - 1\right]^{-\frac{1}{2}}$. Since $T(n)$ is increasing and $P_f < \frac{\sigma_w^2}{2}$, $\frac{U_{T(n)}}{\sqrt{T(n)}} \xrightarrow{\mathcal{P}} 0$ as $n \to \infty$.

To show that $U_{t_A}/\sqrt{V(n)}$ in (12) also converges in probability to zero, we need the first two moments of U_{t_A} defined in (13). We use the MGF $\mathcal{M}_{\chi_n^2}(x) = (1-2x)^{-n/2}$ evaluated at $x = P_f/2\sigma_w^2$ to compute the expectation:

$$\mathbb{E}[U_{t_A}] = \mathbb{E}\left[\exp\left(\frac{P_f X_{t_A}}{2\sigma_w^2}\right)\right] = \left(\frac{\sigma_w^2}{\sigma_w^2 - P_f}\right)^{\frac{n}{2}} \quad (17)$$

The second moment of U_{t_A} is:

$$\mathbb{E}[U_{t_A}^2] = \mathbb{E}\left[\exp\left(\frac{P_f X_{t_A}}{\sigma_w^2}\right)\right] = \left(\frac{\sigma_w^2}{\sigma_w^2 - 2P_f}\right)^{\frac{n}{2}} \quad (18)$$

The probability that the magnitude of the term $\frac{U_{t_A}}{\sqrt{V(n)}}$ in (12) exceeds $\delta > 0$ is upper-bounded using Chebyshev's inequality:

$$\mathbb{P}\left(\left|\frac{U_{t_A}}{\sqrt{V(n)}}\right| > \delta\right) \leq \frac{\mathrm{Var}[U_{t_A}]}{\left(\delta\sqrt{V(n)} - \mathbb{E}[U_{t_A}]\right)^2} \quad (19)$$

Dividing the numerator and denominator in the RHS of (19) by $\mathrm{Var}[U_{t_A}]$, we note that $\mathbb{E}[U_{t_A}]/\sqrt{\mathrm{Var}[U_{t_A}]} = \left(\left(1 + \frac{P_f^2}{\sigma_w^4 - 2P_f\sigma_w^2}\right)^{\frac{n}{2}} - 1\right)^{-\frac{1}{2}} < C$, with C a constant for $P_f < \sigma_w^2/2$. Also, $V(n)/\mathrm{Var}[U_{t_A}] \geq V(n)/\mathbb{E}[U_{t_A}^2]$ and

$$\frac{V(n)}{\mathbb{E}[U_{t_A}^2]} \geq (T(n)-1)\left[\left(1 - \frac{2P_f^2}{\sigma_w^4}\right)^{\frac{n}{2}} - \left(1 - \frac{3P_f^2}{\sigma_w^4}\right)^{\frac{n}{2}}\right]. \quad (20)$$

The term inside the square brackets in (20) is dominated by $\left(1 - \frac{2P_f^2}{\sigma_w^4}\right)^{\frac{n}{2}} = e^{\frac{n}{2}\log\left(1 - \frac{2P_f^2}{\sigma_w^4}\right)}$. Let's set the symbol power to $P_f = \frac{\sigma_w^2\sqrt{\log T(n)}}{2\sqrt{n}}$. Using the Taylor series expansion of $\log(1-x)$ at $x = 0$, we can demonstrate that, when $T(n) = o(e^n)$, $\frac{U_{t_A}}{\sqrt{V(n)}} \xrightarrow{\mathcal{P}} 0$. When $T(n) = \Omega(e^n)$, convergence is obtained with the symbol power set to a constant $P_f < \frac{\sigma_w^2}{2}$.

When $\left|U_{T(n)}/\sqrt{V(n)}\right| < \delta$, the false alarm probability is lower-bounded as follows:

$$\mathbb{P}(L^{(n)} > S(n) | H_0 \text{ is true}) \geq \mathbb{P}\left(E_g(S(n),\delta)\right), \quad (21)$$

where $E_g(S(n),\delta)$ denotes the event that $\frac{1}{\sqrt{V(n)}}\sum_{t=1}^{T(n)-1}(U_t - M(n)) \geq S(n) + \delta$. Similarly, when $\left|U_{t_A}/\sqrt{V(n)}\right| < \delta$, the probability of missed detection is lower-bounded as follows:

$$\mathbb{P}(L^{(n)} \leq S(n) | H_1 \text{ is true}) \geq \mathbb{P}\left(E_l(S(n),\delta)\right), \quad (22)$$

where $E_l(S(n),\delta)$ denotes the event that $\frac{1}{\sqrt{V(n)}}\sum_{t=1,t\neq t_A}^{T(n)}(U_t - M(n)) \leq S(n) - \delta$. Denote by $E_C(S(n),\delta)$ the event when either event $E_g(S(n),\delta)$ occurs when Alice is quiet or event $E_l(S(n),\delta)$ occurs when Alice transmits. Since we assume equiprobable priors,

$$\mathbb{P}(E_C(S(n),\delta)) = \frac{\mathbb{P}\left(E_g(S(n),\delta)\right) + \mathbb{P}\left(E_l(S(n),\delta)\right)}{2}. \quad (23)$$

By the CLT for triangular arrays in [6, Th. 27.2], the normalized sums in the events $E_g(S(n),\delta)$ and $E_l(S(n),\delta)$ converge in distribution to standard Gaussian random variables. This result only provides pointwise convergence in the argument of the distribution function, but $S(n)$ is the n^{th} value in an arbitrary sequence. Instead, in [7, App. A], we exploit the uniform convergence on any finite number of points and the monotonicity of the distribution function to show that, for each normalized sum, setting $\delta = \epsilon\sqrt{2\pi}/9$ yields n_0 such that for all $n \geq n_0$ and any $S(n)$,

$$\mathbb{P}\left(E_C\left(S(n), \frac{\epsilon\sqrt{2\pi}}{9}\right)\right) \geq \frac{1}{2} - \frac{\epsilon}{3}. \quad (24)$$

By (16) and (19), there exists n_1 such that for all $n \geq n_1$, $\mathbb{P}\left(\left|U_{T(n)}/\sqrt{V(n)}\right| > \frac{\epsilon\sqrt{2\pi}}{9}\right) < \frac{\epsilon}{3}$ and $\mathbb{P}\left(\left|U_{t_A}/\sqrt{V(n)}\right| > \frac{\epsilon\sqrt{2\pi}}{9}\right) < \frac{\epsilon}{3}$. The intersection of these events and the event $E_C(S(n), \epsilon\sqrt{2\pi}/9)$ yields an error event. Combining their probabilities using DeMorgan's Law and the union bound, we can lower-bound $\mathbb{P}_e^{(w)} \geq \frac{1}{2} - \epsilon$ for all $n \geq \max\{n_0, n_1\}$, concluding the analysis of Willie's detector.

Analysis (Bob): The analysis of Bob's decoder follows from the same arguments as used in [2]. ■

IV. Converse

Theorem 2. *If Alice attempts to transmit $\omega(\sqrt{n\log T(n)})$ bits on an arbitrary slot of length n symbol periods selected from a set of $T(n)$ slots, then, as $n \to \infty$, either there exists a detector that Willie can use to detect her with high probability, or Bob cannot decode with arbitrarily low probability of error.*

Proof: Let Willie use the following hypothesis test on his observations of the channel from Alice $\mathbf{y}_w$: with $Y_t = \sum_{y_i \in \mathbf{y}_w(t)} y_i^2$ being the observed power in each slot and $Y_{\max} = \max_{t\in\{1,\ldots,T(n)\}} Y_t$, reject or accept the null hypothesis based on a comparison of $Y_{\max}$ with a threshold S.

Suppose Alice does not transmit. Willie's probability of false alarm is $\mathbb{P}(Y_{\max} > S)$. Let $S = \sigma_w^2(n + \sqrt{n}\delta)$. To find δ so that Willie's detector has an arbitrary probability of false alarm $\mathbb{P}_{FA}^*$ as $n \to \infty$, note that each $Y_t = \sigma_w^2 X_t$ where $\{X_t\}$, $X_t \sim \chi_n^2$, $t = 1,\ldots,T(n)$ is a sequence of i.i.d. chi-squared random variables each with n degrees of freedom. We have:

$$\mathbb{P}(Y_{\max} > S) = 1 - \mathbb{P}\left(X_{\max} \leq S/\sigma_w^2\right) \quad (25)$$

$$= 1 - \left(1 - \mathbb{P}\left(X_1 > n + \sqrt{n}\delta\right)\right)^{T(n)} \quad (26)$$

where $X_{\max} = \max_{t\in\{1,\ldots,T\}} X_t$. For the desired $\mathbb{P}_{FA}^*$,

$$1 - (1 - \mathbb{P}_{FA}^*)^{1/T(n)} = \mathbb{P}\left(X_1 > n + \sqrt{n}\delta\right). \quad (27)$$

Using a Chernoff bound for the tail of a chi-squared distribution [8, Lemma 2.2], we obtain:

$$\mathbb{P}\left(X_1 > n + \sqrt{n}\delta\right) \leq \left(1 + \delta/\sqrt{n}\right)^{n/2} e^{-\frac{\sqrt{n}\delta}{2}} \tag{28}$$

$$= e^{\frac{n}{2}\log\left(1+\frac{\delta}{\sqrt{n}}\right) - \frac{\sqrt{n}\delta}{2}} \tag{29}$$

$$= e^{-\delta^2/4 + \mathcal{O}(1/\sqrt{n})} \tag{30}$$

with (30) due to Taylor series expansion of $\log(1+x)$ at $x = 0$. Discarding low order terms and solving (30) for δ yields $\delta = 2\sqrt{-\log\left(1-(1-\mathbb{P}_{FA}^*)^{1/T(n)}\right)}$. Taylor series expansion of $1 - c^x$ at $x = 0$ yields $1-(1-\mathbb{P}_{FA}^*)^{1/T(n)} = \frac{1}{T(n)}\log\left(\frac{1}{1-\mathbb{P}_{FA}^*}\right) + \mathcal{O}(\frac{1}{T^2(n)})$. Thus, setting $\delta = c\sqrt{\log T(n)}$ with some constant $c > 0$ yields the desired probability of false alarm $\mathbb{P}_{FA}^*$.

Now suppose Alice picks slot t_A and uses an arbitrary codebook $\{\mathbf{c}(W_k), k = 1, \ldots, 2^{nR}\}$. Let Alice transmit some codeword $\mathbf{c}(W_k)$ with average per-symbol power $P_f = \frac{\|\mathbf{c}(W_k)\|^2}{n}$. Willie's probability of missing Alice's transmission is

$$\mathbb{P}_{MD}^{(k)} = \mathbb{P}(Y_{\max} \leq S) = \mathbb{P}(Y_{t_A} \leq S) \prod_{\substack{t=1 \\ t \neq t_A}}^{T(n)} \mathbb{P}(Y_t \leq S) \tag{31}$$

where the factorization in (31) is due to the independence of Alice's codeword and the noise in other slots. $\prod_{t=1, t\neq t_A}^{T(n)} \mathbb{P}(Y_t \leq S) \leq 1$ does not depend on Alice's codeword. However, since the codeword is an unknown deterministic signal that is added to AWGN on Willie's channel to Alice, $\frac{Y_{t_A}}{\sigma_w^2} \sim \chi_n^2(nP_f)$ is a non-central chi-squared random variable with n degrees of freedom and non-centrality parameter $\frac{nP_f}{\sigma_w^2}$. The expected value and variance of Y_{t_A} are [9, App. D.1]:

$$\mathbb{E}\left[Y_{t_A}\right] = \sigma_w^2 n + nP_f \tag{32}$$

$$\mathrm{Var}\left[Y_{t_A}\right] = 2n\sigma_w^4 + 4n\sigma_w^2 P_f \tag{33}$$

Chebyshev's inequality with (32) and (33) yields:

$$\mathbb{P}_{MD}^{(k)} \leq \mathbb{P}\left(\left|Y_{t_A} - \sigma_w^2 n - nP_f\right| > nP_f - c\sigma_w^2\sqrt{n\log T(n)}\right)$$

$$\leq \frac{2\sigma_w^4 + 4\sigma_w^2 P_f}{\left(\sqrt{n}P_f - c\sigma_w^2\sqrt{\log T(n)}\right)^2} \tag{34}$$

If $P_f = \omega\left(\sqrt{\frac{\log T(n)}{n}}\right)$, as $n \to \infty$, Willie's average probability of error can be made arbitrarily low.

The proof of the non-zero lower bound on Bob's probability of decoding error if Alice tries to transmit $\omega(\sqrt{n\log T(n)})$ bits using average symbol power $P_f = \mathcal{O}\left(\sqrt{\frac{\log T(n)}{n}}\right)$ follows from a similar proof in [2, Sec. IV]. ∎

V. Relationship with Steganography

Steganographic systems [10] hide information by altering the properties of fixed-size, finite-alphabet covertext objects (e.g. images), and are subject to a similar square root law as LPD communication: $\mathcal{O}(\sqrt{n})$ symbols in covertext of size n may safely be modified to hide an $\mathcal{O}(\sqrt{n}\log n)$-bit message [11]. The similarity between the square root laws in these disciplines is due to the mathematics of statistical hypothesis testing. However, in steganography, the transmission to Bob is noiseless, which allows the extra $\log n$ factor.

Batch steganography uses multiple covertext objects to hide a message and is subject to the steganographic square root law described above [12], [13]. The batch steganography interpretation of LPD communication using timing as described in this work is equivalent to using only one of $T(n)$ covertext objects of size n to embed a message. Willie, who knows that one covertext object is used but not which one, has to examine all of them. We are not aware of any work on this particular problem, but it is likely that our result extends to it.

VI. Conclusion and Future Work

We have shown that secretly pre-arranging a choice of a single n-symbol period slot out of $T(n)$ allows Alice to reliably transmit $\mathcal{O}(\sqrt{n\log T(n)})$ bits on an AWGN channel to Bob while rendering Willie's detector arbitrarily close to ineffective. Surprisingly, the multiplicative increase in transmitted information over the result in [1], [2] is obtained at only an additive cost of $\log T(n)$ pre-shared secret bits.

In the future we plan on extending these results to peak power constrained transmitter. Not only would this make the communication scheme realistic, but also would provide an upper bound on the amount of required pre-shared secret bits.

References

[1] B. A. Bash, D. Goeckel, and D. Towsley, "Square root law for communication with low probability of detection on AWGN channels," in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, Cambridge, MA, Jul. 2012.

[2] ——, "Limits of reliable communication with low probability of detection on AWGN channels," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 1921–1930, 2013, arXiv:1202.6423.

[3] P. H. Che, M. Bakshi, and S. Jaggi, "Reliable deniable communication: Hiding messages in noise," in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, Istanbul, Turkey, Jul. 2013, arXiv:1304.6693.

[4] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd ed. Cambridge, Massachusetts: MIT Press, 2001.

[5] E. Lehmann and J. Romano, *Testing Statistical Hypotheses*, 3rd ed. New York: Springer, 2005.

[6] P. Billingsley, *Probability and Measure*, 3rd ed. New York: Wiley, 1995.

[7] B. A. Bash, D. Goeckel, and D. Towsley, "LPD Communication when the Warden Does Not Know When," arXiv:1403.1013, 2014.

[8] S. Dasgupta and A. Gupta, "An Elementary Proof of a Theorem of Johnson and Lindenstrauss," *Random Struct. Algorithms*, vol. 22, no. 1, pp. 60–65, Jan. 2003.

[9] D. Torrieri, *Principles of Spread-spectrum Communication Systems*. Boston, MA, USA: Springer, 2005.

[10] J. Fridrich, *Steganography in Digital Media: Principles, Algorithms, and Applications*, 1st ed. New York: Cambridge University Press, 2009.

[11] A. D. Ker, "The square root law does not require a linear key," in *Proceedings of the 12th ACM workshop on Multimedia and security*, ser. MM&Sec '10. New York, NY, USA: ACM, 2010, pp. 213–224.

[12] ——, "A capacity result for batch steganography," *IEEE Signal Processing Letters*, vol. 14, no. 8, pp. 525–528, Aug. 2007.

[13] ——, "Batch steganography and pooled steganalysis," in *Information Hiding*, ser. Lecture Notes in Computer Science, J. L. Camenisch, C. S. Collberg, N. F. Johnson, and P. Sallee, Eds. Springer Berlin Heidelberg, 2007, vol. 4437, pp. 265–281.