Pilot-Attacks Can Enable Positive-Rate Covert Communications of Wireless Hardware Trojans

Serhat Bakirtas[1], Matthieu R. Bloch[2], and Elza Erkip[1]

[1]Tandon School of Engineering, New York University, Brooklyn, NY
{serhat.bakirtas, elza}@nyu.edu
[2]School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA
matthieu.bloch@coe.gatech.edu

***Abstract*—Hardware Trojans, by exploiting the link margins inherent in communication systems, can inflict harm on wireless networks. This paper investigates a four-terminal setting where, alongside the legitimate communication link, a hardware Trojan embedded in the legitimate transmitter attempts to establish communication with its intended rogue receiver. In the channel estimation phase, the Trojan carries out a covert pilot scaling attack to corrupt the channel estimation of the legitimate receiver. Subsequently, in the communication phase, the Trojan exploits the ensuing imperfect channel estimation to covertly communicate with its receiver. By analyzing the corresponding hypothesis tests conducted by the legitimate receiver in both phases, we establish that the pilot scaling attack allows the Trojan to operate in the so-called "linear regime" *i.e.,* covertly and reliably transmitting at a positive rate to the rogue receiver. Overall, our results highlight the vulnerability of the channel estimation process in wireless communication systems against hardware Trojans.**

***Index Terms*—hardware Trojans, wireless communications, covert communications, pilot corruption attack**

I. Introduction

In communication networks, the assurance of confidentiality, integrity, and authenticity of transmissions has always been paramount. Recently, however, the concept of achieving a low probability of detection, or *covertness*, has garnered increased attention [1]. This renewed interest is partly driven by the understanding that the mere knowledge of a party's communication can be as significant as the content of the communication itself. It also stems from concerns about potential side channels that could surreptitiously exfiltrate sensitive information [2]. This work is particularly motivated by the latter concern, focusing on the opportunities that hardware Trojans have to exist "in the margins". These margins are inherent in communication protocols, designed to accommodate minor imperfections [3], [4]. Given the ubiquity of pilot symbols in contemporary wireless protocols, this study aims to explore the feasibility and impact of attacks where hardware Trojans manipulate these pilot symbols in an undetected way. The ultimate goal is to understand how such manipulation could undermine the detection capabilities of monitoring entities in subsequent transmissions.

Theoretical explorations of covert capacity have unveiled two distinct regimes of covert communications. The first is the *square-root law* regime [1], [5], [6], where the number of covert bits must scale with the square root of the blocklength. The second is the *linear* regime, where the number of bits can scale linearly with the blocklength [7]. Operating within the linear regime typically necessitates the exploitation of uncertainty in channel state knowledge [8]–[11]. Specifically, the introduction of artificial noise is a potent signaling technique used to engineer this uncertainty [12].

In this study, we examine a scenario where a hardware Trojan manipulates pilot symbols with the intent to diminish the channel estimation accuracy of legitimate parties. This manipulation subsequently curtails their capacity to detect communication initiated by the hardware Trojan. A significant contribution of our research is the demonstration of how pilot symbol manipulation by a hardware Trojan can effectively bypass the square root law, thereby facilitating operation within the linear regime. This finding underscores the potential risks posed by hardware Trojans in modern communication systems.

The organization of the rest of this paper is as follows. In Section II we formally introduce the system model. In Section III, we present our main results and their proof. Finally, in Section IV, we offer concluding remarks and discuss future directions. The full proofs are provided in the longer version of this paper [13].

Notation: We denote scalars with lowercase letters, vectors with lowercase bold letters, and matrices with uppercase bold letters. For vectors $\|.\|_2$ denotes the Euclidian norm and for matrices $|.|$ denotes the determinant. $\mathcal{CN}(\mu,\sigma^2)$ denotes circularly-symmetric complex Gaussian distribution with respective mean μ and variance σ^2. $\mathbb{D}$ denotes the Kullback-Leibler divergence [14, Chapter 2.3]. $\mathcal{O}$, o, Θ, and ω denote the standard Bachmann–Landau notations [15, Chapter 3]. Unless stated otherwise, log denotes the natural logarithm. Finally, L and n denote the pilot sequence length and communication block length, respectively, and for a sequence $\{r_i\}_{i=1}^{n+L}$, we refer to $\{r_i\}_{i=1}^{L}$ and $\{r_i\}_{i=L+1}^{n+L}$ by $\boldsymbol{r}^{\text{est}}$ and $\boldsymbol{r}^{\text{comm}}$, indicating channel estimation and communication phases, respectively.

This work is supported by National Science Foundation grants 1815821 and 2148293.

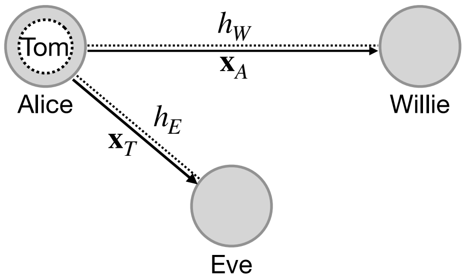

Fig. 1. The legitimate transmitter, Alice, communicates with her intended (legitimate) receiver, Willie. Simultaneously, the hardware Trojan, Tom, embedded in Alice, also communicates with his intended rogue receiver, Eve. Willie's objective is to decode Alice's signal $\boldsymbol{x}_A$ and detect the existence of any rogue signal $\boldsymbol{x}_T$.

II. Problem Formulation

As illustrated in Figure 1, we consider a four-terminal scenario in which (1) Alice, legitimate transmitter, communicates with Willie, legitimate receiver; (2) Tom, hardware Trojan embedded in Alice, seeks to simultaneously communicate with Eve, rogue receiver, while evading detection by Willie, who also acts as the warden.

We assume that Alice, Tom, Willie, and Eve each have a single antenna. We adopt a Rayleigh block fading channel model by which the received signal $\boldsymbol{y}_u$ at the user $u \in \{W, E\}$ is given by

$$y_{u,i} = \alpha_u \, h_u \, x_i \, + \, z_{u,i}, \quad \forall i \in [n+L] \tag{1}$$

where α_W and h_W (resp. α_E and h_E) denote the propagation loss and the channel fading gain between Alice and Willie (resp. Tom and Eve). We assume here that $h_u \sim \mathcal{CN}(0, \sigma_H^2)$ is independent of the transmitted sequence $\{x_i\}_{i=1}^{n+L}$ and noise $\{z_{u,i}\}_{i=1}^{n+L}$ and remains constant for at least $n+L$ symbol periods. We assume $z_{u,i} \overset{\text{i.i.d.}}{\sim} \mathcal{CN}(0, \sigma_u^2)$.

Since Tom is embedded in Alice, h_W and α_W (resp. h_E and α_E) also denote the channel gain and propagation loss between Tom and Willie (resp. Alice and Eve).

Motivated by practical wireless communication networks, our proposed system model comprises two distinct phases: In the first phase, called the *channel estimation phase*, Alice sends a known pilot sequence for Willie to estimate the channel. Meanwhile, in an effort to improve his chances at covertness in the subsequent phase, Tom tries to covertly corrupt Willie's estimate by scaling the pilot sequence by $1+\varepsilon$ for some small $\varepsilon > 0$ called the *scaling parameter*. Simultaneously, Willie attempts to detect whether the pilot sequence is corrupted by scaling or not. In the second phase, called the *communication phase*, Alice and Tom communicate with their respective receivers Willie and Eve, while Willie once again attempts to detect any rogue communication between Tom and Eve.

Willie's detection falls under the framework of simple binary hypothesis tests. We let H_0 denote the null hypothesis in the channel estimation phase when Alice's pilot sequence is not corrupted by Tom and H_1 denote the alternative hypothesis when Alice's pilot sequence is scaled by $(1+\varepsilon)$ by Tom. Formally, we have

$$\begin{array}{lll} H_0: & x_i = s_{A,i} & \forall i \in [L] \\ H_1: & x_i = (1+\varepsilon) s_{A,i}, & \forall i \in [L] \end{array} \tag{2}$$

where $\boldsymbol{s}_A$ is Alice's pilot sequence of length L, known to Tom, Eve and Willie. We assume $L = o(n)$ and $\|\boldsymbol{s}_A\|_2^2 = \omega_L(1)$ to enable reliable channel estimation [16, Section III-A].

We let $\tilde{H}_0$ denote the null hypothesis in the communication phase when the only transmission is between Alice and Willie, and $\tilde{H}_1$ denote the alternative hypothesis where in addition to the legitimate communication link between Alice and Willie, Tom also communicates with Eve. More formally, we have

$$\begin{array}{lll} \tilde{H}_0: & x_{L+i} = x_{A,i}, & \forall i \in [n] \\ \tilde{H}_1: & x_{L+i} = x_{A,i} + x_{T,i}, & \forall i \in [n] \end{array} \tag{3}$$

where $\boldsymbol{x}_A$ and $\boldsymbol{x}_T$ denote Alice's and Tom's channel inputs, respectively.

We assume that $\boldsymbol{x}_A$ and $\boldsymbol{x}_T$ are mutually orthogonal zero-mean complex Gaussian sequences with *i.i.d.* components with a deterministic short-term [17] power constraint. More formally,

$$\frac{1}{n}\|\boldsymbol{x}_A\|_2^2 = \Lambda_A, \qquad \frac{1}{n}\|\boldsymbol{x}_T\|_2^2 = \Lambda_T \tag{4}$$

In addition, we assume that $\boldsymbol{x}_A$ and $\boldsymbol{x}_T$ are orthogonal to $\boldsymbol{z}^{\text{comm}}$. This is justified by independence in the asymptotic $(n \to \infty)$ regime.

We further assume that both Alice and Tom know h_W and h_E perfectly. This, for example, could happen in a TDD system with channel reciprocity in which Eve is also a legitimate user and no Trojan is present at Willie or Eve to cause pilot corruption.

As argued in [3], [4], the transmitters in typical wireless communication scenarios do not operate at the capacity because of design choices. We, therefore, assume that Alice adopts a link margin, transmitting at a rate strictly lower than the instantaneous channel capacity to Willie for given channel realization h_W under a short-term power constraint. More formally, we assume that Alice transmits to Willie at a rate R_A such that

$$R_A < \log_2\left(1 + \frac{\alpha_W^2 |h_W|^2 \Lambda_A}{\sigma_W^2}\right). \tag{5}$$

Throughout, we assume Alice and Tom use their knowledge of h_W and h_E to ensure no outage takes place in their respective communications. Hence, the instantaneous capacity from Alice to Willie given by the RHS of Eq. (5), also known as the delay-limited capacity [18], is the appropriate bound for R_A.

Observe that Willie can always perform a *blind test* ignoring his received signal and pick hypotheses based on an independent Bernoulli random variable, still achieving $\mathbb{P}_F + \mathbb{P}_M = 1$ in either phase, where $\mathbb{P}_F$ and $\mathbb{P}_M$ denote the false alarm and missed detection probabilities. Hence, as is customary in the literature [1], [5], [19], [20], Tom's covertness objective is to make Willie's detection strategy comparable to a blind test.

For tractability, we assume that Willie performs the hypothesis tests in different phases separately. More specifically, if Willie's test does not perform credibly better than a blind test in the channel estimation phase (Eq. (6)), Willie fails to detect Tom and acts based on the null hypothesis H_0 in the communication phase. Thus, Tom's subsequent objective becomes preying on Willie's initial failure and communicating covertly to Eve (Eq. (7)). While doing so, Tom's actions should not disrupt successful decoding in the legitimate Alice-Willie link (Eq. (8)). We note that the communication phase commences only if Willie fails to detect the Trojan in the channel estimation phase.

Our covertness criteria described above are formally defined in Definition 1 below.

Definition 1. (Covertness Criteria) Given a *detection budget* (δ_1, δ_2) and Alice's transmit power and rate pair (Λ_A, R_A), Tom is said to remain *covert* if

$$\lim_{L\to\infty} \mathbb{P}_F^{(1)} + \mathbb{P}_M^{(1)} \geq 1 - \delta_1 \tag{6}$$

$$\lim_{n\to\infty} \mathbb{P}_F^{(2)} + \mathbb{P}_M^{(2)} \geq 1 - \delta_2 \tag{7}$$

$$\lim_{n\to\infty} P_{\text{error}}^{(n)} = 0, \tag{8}$$

where

$$\mathbb{P}_F^{(1)} \triangleq \Pr(\hat{H}_e = H_1 | H_0) \tag{9}$$

$$\mathbb{P}_M^{(1)} \triangleq \Pr(\hat{H}_e = H_0 | H_1) \tag{10}$$

$$\mathbb{P}_F^{(2)} \triangleq \Pr(\hat{H}_c = \tilde{H}_1 | \tilde{H}_0, H_1, \hat{H}_e = H_0) \tag{11}$$

$$\mathbb{P}_M^{(2)} \triangleq \Pr(\hat{H}_c = \tilde{H}_0 | \tilde{H}_1, H_1, \hat{H}_e = H_0) \tag{12}$$

Here $\hat{H}_e$ and $\hat{H}_c$ denote Willie's decision in the channel estimation and the communication phases, respectively, and $P_{\text{error}}^{(n)}$ denotes the probability that Willie cannot decode $\boldsymbol{x}_A$. We have $\delta_1 = \Theta_L(1)$ and $\delta_2 = \Theta_n(1)$ are small but non-vanishing constants.

In this paper, our main objective is to find whether for given $(\delta_1, \delta_2, \Lambda_A, R_A)$, Tom can drive the system to the linear regime, *i.e.*, communicate with Eve covertly at a non-zero rate by a proper choice of ε and Λ_T, and, if so, to study this covert rate.

III. Main Results

We now state and prove our main results regarding the achievability of positive covert rate.

We first consider a positive detection budget, *i.e.* $\delta_1, \delta_2 > 0$. Our main result in Theorem 1 is an achievable set of covert rates.

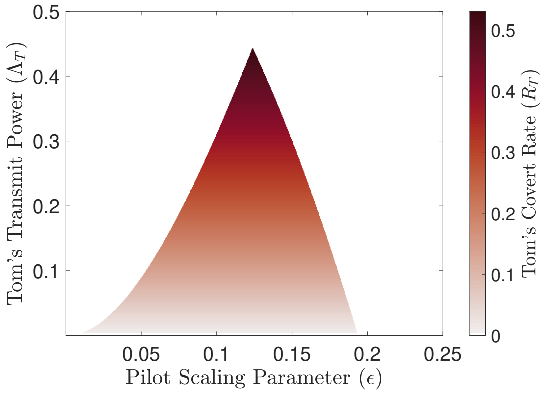

Fig. 2. A heatmap demonstrating the relationship between Tom's pilot scaling parameter ε, his transmit power Λ_T, and the achievable covert rate R_T given in Theorem 1, where colors indicate the values of R_T across the range of ε and Λ_T when Eve can perform interference cancellation (See (19)). Here, $\alpha_W^2 = \alpha_E^2 = 0.1$, $|h_W|^2 = |h_E|^2 = 1$, $\sigma_W^2 = \sigma_E^2 = 0.1$, $\delta_1 = 1/\sqrt{10}$, and $\Lambda_A = 20$ where Alice transmits at ($\approx$ 3.5 bpcu) 80% of her capacity ($\approx$ 4.4 bpcu) to Willie (See (5)).

Theorem 1. ***(Achievable Covert Rate when $\delta_1 > 0$)*** *Consider a* detection budget (δ_1, δ_2) *with* $\delta_1 > 0$ *and* $\delta_2 \in (0, 1)$, *and Alice's transmit power and rate pair* (Λ_A, R_A). *If Tom's scaling parameter* ε *and transmit power* Λ_T *satisfy*

$$\varepsilon \leq \frac{\delta_1}{\sqrt{2}} \tag{13}$$

$$\tau^* < \varepsilon^2 \alpha_W^2 |h_W|^2 \Lambda_A + \sigma_W^2 \tag{14}$$

$$R_A \leq \log_2(1 + \gamma_W) \tag{15}$$

where

$$\tau^* = (1+\varepsilon)^2 \alpha_W^2 |h_W|^2 \Lambda_T \frac{\exp\left(\frac{(1+\varepsilon)^2 \alpha_W^2 |h_W|^2 \Lambda_T}{\sigma_W^2}\right)}{\exp\left(\frac{(1+\varepsilon)^2 \alpha_W^2 |h_W|^2 \Lambda_T}{\sigma_W^2}\right) - 1} \tag{16}$$

$$\gamma_W = \frac{\alpha_W^2 |h_W|^2 \Lambda_A}{\varepsilon^2 \alpha_W^2 |h_W|^2 \Lambda_A + \alpha_W^2 |h_W|^2 \Lambda_T + \sigma_W^2} \tag{17}$$

then Tom can communicate with Eve covertly at any rate R_T *satisfying*

$$R_T \leq \log_2\left(1 + \frac{\alpha_E^2 |h_E|^2 \Lambda_T}{\alpha_E^2 |h_E|^2 \Lambda_A + \sigma_E^2}\right). \tag{18}$$

Additionally, if

$$R_A \leq \log_2\left(1 + \frac{\alpha_E^2 |h_E|^2 \Lambda_A}{\alpha_E^2 |h_E|^2 \Lambda_T + \sigma_E^2}\right) \tag{19}$$

then, Tom's rate R_T *can be improved to*

$$R_T \leq \log_2\left(1 + \frac{\alpha_E^2 |h_E|^2 \Lambda_T}{\sigma_E^2}\right). \tag{20}$$

Theorem 1 states that as long as $\delta_1 > 0$, for any $\delta_2 \in (0, 1)$ Tom can communicate with Eve at a positive covert rate, effectively operating in the linear covert regime.

As discussed in detail in Sections III-A through III-D, the conditions in Theorem 1 have the following interpretation: Eq. (13) corresponds to Tom satisfying the covertness criterion (6) in the channel estimation phase. Subsequently, Eq. (14) allows Tom to force Willie's test to decide in favor of $\tilde{H}_1$ under both $\tilde{H}_0$ and $\tilde{H}_1$, leading to a blind test, satisfying (7). Eq. (15) ensures that Tom's actions do not disrupt communication over the legitimate link, satisfying (8). Note that to achieve R_T in (18), Eve treats Alice's signal as noise. Finally, the additional constraint (19) allows Eve to perform interference cancellation where she can decode and cancel the legitimate signal $\boldsymbol{x}_A$ by first treating $\boldsymbol{x}_T$ as noise, leading to the covert rate R_T in (20).

Figure 2 illustrates the (ε, Λ_T) pairs satisfying Theorem 1 and the corresponding achievable positive covert rates R_T for a given configuration. A key observation is that the covert rate is not necessarily maximized when Tom maximizes ε, *i.e.*, the channel estimation error of Willie. This is due to Willie's signal-to-interference-plus-noise ratio (SINR) degradation caused by both mismatched decoding due to his imperfect channel estimation and interference by Tom (See (17)).

Next, we consider a zero detection budget in the channel estimation phase. Our main result regarding the achievable covert rate when $\delta_1 = 0$ is presented in Theorem 2 below.

Theorem 2. ***(Achievable Covert Rate when*** $\boldsymbol{\delta_1 = 0}$***)*** *When* $\delta_1 = 0$*, Tom can transmit covertly if and only if*

$$R_T = \mathcal{O}(n^{-1/2}) \tag{21}$$

for any $\delta_2 \in (0,1)$*.*

Theorem 2 states that when $\delta_1 = 0$, Tom's scaling parameter ε needs to vanish with L and hence Tom cannot conduct an effective pilot scaling attack and in turn, he needs to obey the square root law [1].

The rest of this section details the proof of Theorems 1 and 2. First, we focus on the channel estimation phase in Section III-A and we investigate the covertness of Tom's pilot scaling attack as well as its impact on Willie's channel estimate. Next, we consider the communication phase in Section III-B and discuss Willie's detection strategy and his optimal threshold. In Section III-C, we derive sufficient and necessary conditions under which Tom could transmit at a non-vanishing power Λ_T while remaining covert. In Section III-D, we investigate the SINR deterioration at Willie due to Tom's actions. Finally, in Sections III-E and III-F we present the complete proofs of Theorems 1 and 2.

A. Covert Pilot Scaling & Willie's Channel Estimation Error

We start by deriving sufficient conditions on Tom's scaling parameter ε such that the covertness criterion (6) provided in Definition 1 is met.

Lemma 1. ***(Covert Pilot Scaling)*** *Tom's pilot scaling attack remains covert, for any* $\varepsilon \leq {}^{\delta_1}/\sqrt{2}$*.*

Proof (Sketch). As is customary, we start by observing

$$\mathbb{P}_F^{(1)} + \mathbb{P}_M^{(1)} \geq 1 - \sqrt{\mathbb{D}(\mathbb{P}_1 \| \mathbb{P}_0)} \tag{22}$$

where $\mathbb{D}(\mathbb{P}_1 \| \mathbb{P}_0)$ denotes the Kullback–Leibler divergence between the alternative $\mathbb{P}_1$ and the null $\mathbb{P}_0$ distributions induced by $\boldsymbol{y}_W^{\mathrm{est}}$ under both hypotheses. We obtain

$$\lim_{L \to \infty} \mathbb{D}(\mathbb{P}_1 \| \mathbb{P}_0) = 2\log(1+\varepsilon) - 1 + (1+\varepsilon)^{-2} \tag{23}$$

and argue that $\lim_{L \to \infty} \mathbb{D}(\mathbb{P}_1 \| \mathbb{P}_0) \leq 2\varepsilon^2$. □

For $\varepsilon \leq {}^{\delta_1}/\sqrt{2}$, Willie estimates h_W assuming the null hypothesis H_0 is true. As customary, we assume Willie uses the minimum mean square error (MMSE) estimator $\hat{h}_W$ of h_W.

Proposition 1. ***(Willie's Estimate)*** *For* $\varepsilon \leq {}^{\delta_1}/\sqrt{2}$*, Willie's estimate* $\hat{h}_W \triangleq \lim_{L \to \infty} \hat{h}_W^{mmse}$ *under* H_0 *and* H_1 *is given in*

$$\begin{aligned} H_0 &: \quad \hat{h}_W = h_W \\ H_1 &: \quad \hat{h}_W = (1+\varepsilon) h_W \end{aligned} \tag{24}$$

B. Willie's Detection in the Communication Phase

In this subsection, assuming Willie fails to detect pilot scaling in the channel estimation phase (Eq. (6)),*i.e.*, $\hat{H}_e = H_0$, we discuss his detection strategy for Tom in the communication phase. Throughout, we assume Willie can decode $\boldsymbol{x}_A$ as in (8).

Note that, from (1) and (3), the received signal at Willie is given by

$$\begin{aligned} \tilde{H}_0 &: \quad \boldsymbol{y}_W^{\mathrm{comm}} = \alpha_W \, h_W \, \boldsymbol{x}_A + \boldsymbol{z}_W^{\mathrm{comm}} \\ \tilde{H}_1 &: \quad \boldsymbol{y}_W^{\mathrm{comm}} = \alpha_W \, h_W \, \boldsymbol{x}_A + \alpha_W \, h_W \, \boldsymbol{x}_T + \boldsymbol{z}_W^{\mathrm{comm}} \end{aligned} \tag{25}$$

In Lemma 2 below, we show that Willie's optimal detection strategy is to adopt a radiometer, similar to [19], [21].

Lemma 2. ***(Willie's Optimal Detection Strategy)*** *Given that Willie can decode* $\boldsymbol{x}_A$ *correctly, under* H_0 *Willie's optimal detection strategy is given by*

$$T(\boldsymbol{y}_W^{comm}) \triangleq \frac{1}{n} \| \boldsymbol{y}_W^{comm} - \alpha_W \hat{h}_W \boldsymbol{x}_A \|_2^2 \underset{\tilde{H}_0}{\overset{\tilde{H}_1}{\gtrless}} \tau^\dagger \tag{26}$$

where

$$\tau^\dagger \triangleq \alpha_W^2 |\hat{h}_W|^2 \Lambda_T \frac{e^{\frac{n}{n-1} \frac{\alpha_W^2 |\hat{h}_W|^2 \Lambda_T}{\sigma_W^2}}}{e^{\frac{n}{n-1} \frac{\alpha_W^2 |\hat{h}_W|^2 \Lambda_T}{\sigma_W^2}} - 1} \tag{27}$$

and $\hat{h}_W$ *is Willie's estimate of* h_W*.*

Proof (Sketch). The test statistic $T(\boldsymbol{y}_W^{\mathrm{comm}})$ is found by computing and simplifying the log-likelihood ratio based on $\boldsymbol{y}_W^{\mathrm{comm}}$. Then, the optimal threshold is given by

$$\tau^\dagger \triangleq \operatorname*{argmin}_{\tau > 0} \mathbb{P}_F^\dagger + \mathbb{P}_M^\dagger \tag{28}$$

where

$$\mathbb{P}_F^\dagger \triangleq \Pr(T(\boldsymbol{y}_W^{\mathrm{comm}}) > \tau | \tilde{H}_0, H_0) \tag{29}$$

$$\mathbb{P}_M^\dagger \triangleq \Pr(T(\boldsymbol{y}_W^{\mathrm{comm}}) < \tau | \tilde{H}_1, H_0) \tag{30}$$

Solving $\left.\frac{\partial(\mathbb{P}_F^\dagger + \mathbb{P}_M^\dagger)}{\partial \tau}\right|_{\tau = \tau^\dagger} = 0$, we obtain (27). □

Observe that a key difference between our system model and those of [19], [21] is the existence of the legitimate signaling by Alice. Furthermore, Lemma 2 suggests that before using the radiometer, Willie decodes and cancels the legitimate signal $\boldsymbol{x}_A$. In fact, under H_0, Willie's channel estimate $\hat{h}_W$ is perfect and hence our test statistic $T(\boldsymbol{y}_W^{\text{comm}})$ corresponds to that of [19], [21].

Since Willie is unaware of pilot scaling by ε and in turn believes his cancellation of $\mathbf{x}_A$ has been perfect, his test threshold $\tau^\dagger$ is independent of Λ_A. Furthermore, one can easily verify that $\tau^\dagger$ is increasing in n.

C. Covertly Transmitting At A Positive Power

When Tom performs a pilot scaling attack and Willie cannot detect it, *i.e.,* $\hat{H}_e = H_0$ while H_1 is true, using Proposition 1, Willie's test statistic $T(\boldsymbol{y}_W^{\text{comm}})$ becomes

$$\begin{aligned} \tilde{H}_0 &: \ T(\boldsymbol{y}_W^{\text{comm}}) = \varepsilon^2 \alpha_W^2 |h_W|^2 \Lambda_A + \Lambda_Z \\ \tilde{H}_1 &: \ T(\boldsymbol{y}_W^{\text{comm}}) = \varepsilon^2 \alpha_W^2 |h_W|^2 \Lambda_A + \alpha_W^2 |h_W|^2 \Lambda_T + \Lambda_Z \end{aligned} \tag{31}$$

where $\Lambda_Z \triangleq \frac{1}{n} \|\boldsymbol{z}_W^{\text{comm}}\|_2^2$. Note that due to the imperfect cancellation of the legitimate signal $\boldsymbol{x}_A$, in (31) (See also (24)) there is a residual term depending on ε and Λ_A under both hypotheses.

Under H_1, we have $\tau^* = \lim\limits_{n\to\infty} \tau^\dagger$, where τ^* is given in (16). When Willie decides H_0, for sufficiently large n, his test will be

$$T(\boldsymbol{y}_W^{\text{comm}}) \underset{\tilde{H}_0}{\overset{\tilde{H}_1}{\gtrless}} \tau^*. \tag{32}$$

Now we analyze the performance of Willie's test under H_1 in terms of $\mathbb{P}_F^{(2)}$ and $\mathbb{P}_M^{(2)}$.

Lemma 3. ***(Covert Communications With Pilot Scaling)*** *As long as Willie's optimal threshold satisfies*

$$\tau^* < \varepsilon^2 \alpha_W^2 |h_W|^2 \Lambda_A + \sigma_W^2, \tag{33}$$

or

$$\tau^* > \alpha_W^2 |h_W|^2 (\varepsilon^2 \Lambda_A + \Lambda_T) + \sigma_W^2, \tag{34}$$

we have

$$\lim_{n\to\infty} \mathbb{P}_F^{(2)} + \mathbb{P}_M^{(2)} = 1. \tag{35}$$

Conversely, if

$$\varepsilon^2 \alpha_W^2 |h_W|^2 \Lambda_A + \sigma_W^2 < \tau^* < \alpha_W^2 |h_W|^2 (\varepsilon^2 \Lambda_A + \Lambda_T) + \sigma_W^2, \tag{36}$$

we have

$$\lim_{n\to\infty} \mathbb{P}_F^{(2)} + \mathbb{P}_M^{(2)} = 0. \tag{37}$$

Proof (Sketch). The proof follows from the observation that $\frac{2}{\sigma_W^2} \frac{1}{n} \|\boldsymbol{z}_W^{\text{comm}}\|_2^2 \sim \chi^2(2n)$ and the subsequent use of the tail bounds [22, Lemma 1] for the χ^2-distribution on $T(\boldsymbol{y}_W^{\text{comm}})$ (See Eq. (31)) in order to bound $\mathbb{P}_F^{(2)}$, $\mathbb{P}_M^{(2)}$, $1-\mathbb{P}_F^{(2)}$, or $1-\mathbb{P}_M^{(2)}$ from above, depending on the assumptions on τ^*. □

Observe that when H_1 is true and $\hat{H}_e = H_0$, as $n \to \infty$ we have

$$\begin{aligned} \tilde{H}_0 &: \quad T(\boldsymbol{y}_W^{\text{comm}}) \xrightarrow{p} \varepsilon^2 \alpha_W^2 |h_W|^2 \Lambda_A + \sigma_W^2 \\ \tilde{H}_1 &: \quad T(\boldsymbol{y}_W^{\text{comm}}) \xrightarrow{p} \alpha_W^2 |h_W|^2 (\varepsilon^2 \Lambda_A + \Lambda_T) + \sigma_W^2 \end{aligned} \tag{38}$$

Hence, Lemma 3 states that only if Tom can adjust Λ_T and ε such that Willie sets his test threshold τ^* either below the limit values of $T(\boldsymbol{y}_W^{\text{comm}})$ under both $\tilde{H}_0$ and $\tilde{H}_1$, or above both of these limit values, he can covertly transmit to Eve.

Furthermore, Lemma 3 implies that given $\varepsilon > 0$, as Alice's transmit power Λ_A increases, Tom's chances at covertness improve as the residual term in (31) due to imperfect channel estimation also increases with Λ_A.

Recall that the optimal threshold τ^* is an increasing function of both ε and Λ_T, and the RHS of (33) is a function of ε and Λ_A. Hence, there exists $\Lambda_T^* = \Theta_n(1)$ such that

$$\tau^*|_{\Lambda_T = \Lambda_T^*} = \varepsilon^2 \alpha_W^2 |h_W|^2 \Lambda_A + \sigma_W^2. \tag{39}$$

Therefore, Lemma 3 implies that based on Λ_A and ε, Willie can covertly transmit at a non-vanishing power $\Lambda_T < \Lambda_T^*$.

Conversely, we argue that when there is no pilot scaling attack, *i.e.,* when $\varepsilon = 0$, Tom cannot transmit at a non-vanishing power Λ_T covertly. More specifically, in Lemma 4 below we provide the sufficient and necessary condition on Λ_T for Tom to remain covert.

Lemma 4. ***(Covert Communications with No Pilot Scaling)*** *When $\varepsilon = 0$, Tom can only transmit covertly when he transmits at a power $\Lambda_T = \mathcal{O}(n^{-1/2})$.*

Proof (Sketch). The proof starts with proving that when $\varepsilon = 0$, τ^* satisfies (36) for any $\Lambda_T = \Theta_n(1)$. Hence, by Lemma 3, Tom cannot transmit covertly at any non-vanishing power $\Lambda_T = \Theta_n(1)$. Therefore, to remain covert, Tom needs to transmit at a power $\Lambda_T = o_n(1)$. Next, we assume $\Lambda_T = \mathcal{O}(n^{-1/2})$ and prove the achievability part via an upper bound on $1 - \mathbb{P}_F^{(2)} - \mathbb{P}_M^{(2)}$ by a right Riemann sum and further bounding this Riemann sum via Stirling's approximation [15, Chapter 3.2]. Finally, for the converse result, we assume $\Lambda_T = \omega(n^{-1/2})$ and use tail bounds for χ^2-distribution to show that $\mathbb{P}_F^{(2)} + \mathbb{P}_M^{(2)} \to 0$ as $n \to \infty$. □

D. Willie's SINR Degradation

We stress that when Tom conducts a pilot scaling attack with some $\varepsilon > 0$ and subsequently transmits at a non-vanishing power Λ_T, Willie's SINR deteriorates. More formally, Willie's SINR will become

$$\gamma_W = \frac{\alpha_W^2 |h_W|^2 \Lambda_A}{\varepsilon^2 \alpha_W^2 |h_W|^2 \Lambda_A + \alpha_W^2 |h_W|^2 \Lambda_T + \sigma_W^2} \tag{40}$$

Note that in (40) the first interference term stems from the mismatched decoding due to Tom's pilot scaling attack and in turn Willie's imperfect channel estimation, while the second is due to Tom's transmission.

Furthermore, if ε and Λ_T are large enough such that

$$R_A > \log_2(1 + \gamma_W), \tag{41}$$

Willie will start having decoding errors. Since this unexpected decoding error will imply the existence of a rogue communication, Tom needs to avoid it.

The second constraint (34) in Lemma 3 requires Λ_T to be comparable to or much larger than Λ_A. Since this will disrupt the legitimate Alice-Willie link and in turn be detected, we only focus on satisfying (33).

E. Proof of Theorem 1

We now prove Theorem 1. Observe that Lemma 1 states that as long as (13) is satisfied, the first covertness criterion (6) is satisfied. Hence given (13), Willie estimates $\hat{H}_e = H_0$ and performs channel estimation based on H_0 (See Proposition 1).

Since Tom is successful in the channel estimation phase, Willie is unaware of the pilot scaling attack and conducts his optimal detection strategy in the communication phase as described in Proposition 2. Noticing that $\tau^* = \lim_{n\to\infty} \tau^\dagger$, with τ^* given (16), Lemma 3 states that the second covertness criterion (7) is satisfied.

Finally, as described in Section III-D, (15) ensures that the Alice-Willie link is not disrupted, hence satisfying the final covertness criterion (8) is satisfied. Therefore, given (13)-(15), Tom communicates covertly with Eve who treats $\boldsymbol{x}_A$ as noise, yielding the achievable rate of (18).

As stated in Section III, if (19) is satisfied, Eve first successfully decodes and cancels $\boldsymbol{x}_A$, treating $\boldsymbol{x}_T$ as noise. Note that Eve is aware of Tom's pilot scaling attack and hence her channel estimation and her subsequent cancellation of $\boldsymbol{x}_A$ are perfect. Thus, we obtain the improved rate given in (20). □

F. Proof of Theorem 2

Next, we prove Theorem 2. Begin by observing that, $\delta_1 = 0$ necessitates $\varepsilon = o_L(1)$. Hence by Proposition 1, $\hat{h}_W = h_W$ under both H_0 and H_1. In other words, for any $\varepsilon = o_L(1)$, Tom's pilot scaling attack has no impact on Willie's channel estimation process. Thus, we can only focus on the $\varepsilon = 0$ case.

Next, note that when $\varepsilon = 0$ by Lemma 4, Tom can communicate covertly with Eve only when $\Lambda_T = \mathcal{O}(n^{-1/2})$.

Finally, by performing the MacLaurin series expansion of the subsequent achievable rate with respect to Λ_T, we conclude that when $\varepsilon = o_L(1)$, Tom can covertly communicate with Eve at a rate R_T if and only if $R_T = \mathcal{O}(n^{-1/2})$. □

IV. Conclusion

In this paper, we have investigated a covert communications scenario in which a hardware Trojan carries out a pilot scaling attack to degrade the channel estimate of legitimate parties and subsequently reduces their ability to detect the Trojan's communication. We have showed that for any positive pilot detection budget, the Trojan can effectively drive the system to the linear regime, allowing non-zero covert communication rates. Conversely, we have showed that in the zero pilot detection budget case, the Trojan loses its ability to covertly and effectively corrupt the channel estimation process and in turn has to obey the square root law. Overall, our findings suggest that effective strategies against hardware Trojans also need to take into account the channel estimation phase.

References

[1] B. Bash, D. Goeckel, and D. Towsley, "Limits of reliable communication with low probability of detection on AWGN channels," *IEEE JSAC*, vol. 31, no. 9, pp. 1921–1930, September 2013.

[2] S. Sangodoyin, F. T. Werner, B. B. Yilmaz, C.-L. Cheng, E. M. Ugurlu, N. Sehatbakhsh, M. Prvulovic, and A. Zajic, "Side-channel propagation measurements and modeling for hardware security in IoT devices," *IEEE TAP*, vol. 69, no. 6, pp. 3470–3484, Jun. 2021.

[3] K. S. Subramani, A. Antonopoulos, A. A. Abotabl, A. Nosratinia, and Y. Makris, "Demonstrating and mitigating the risk of an FEC-based hardware Trojan in wireless networks," *IEEE TIFS*, vol. 14, no. 10, pp. 2720–2734, 2019.

[4] K. S. Subramani, N. Helal, A. Antonopoulos, A. Nosratinia, and Y. Makris, "Amplitude-modulating analog/RF hardware Trojans in wireless networks: Risks and remedies," *IEEE TIFS*, vol. 15, pp. 3497–3510, 2020.

[5] M. R. Bloch, "Covert communication over noisy channels: A resolvability perspective," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2334–2354, 2016.

[6] L. Wang, G. W. Wornell, and L. Zheng, "Fundamental limits of communication with low probability of detection," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3493–3503, Jun. 2016.

[7] S.-H. Lee, L. Wang, A. Khisti, and G. W. Wornell, "Covert communication with channel-state information at the transmitter," *IEEE TIFS*, vol. 13, no. 9, pp. 2310–2319, 2018.

[8] P. H. Che, M. Bakshi, C. Chan, and S. Jaggi, "Reliable deniable communication with channel uncertainty," in *2014 IEEE Information Theory Workshop (ITW)*, Hobart, Tasmania, November 2014, pp. 30–34.

[9] T. V. Sobers, B. A. Bash, S. Guha, D. Towsley, and D. Goeckel, "Covert communication in the presence of an uninformed jammer," *IEEE TWC*, vol. 16, no. 9, pp. 6193–6206, 2017.

[10] S. Lee, R. Baxley, M. Weitnauer, and B. Walkenhorst, "Achieving undetectable communication," *IEEE JSTSP*, vol. 9, no. 7, pp. 1195–1205, Oct 2015.

[11] H. Zivari-Fard, M. Bloch, and A. Nosratinia, "Keyless covert communication via channel state information," *IEEE Trans. Inf. Theory*, vol. 68, no. 8, pp. 5440–5474, Aug. 2022.

[12] E. Tekin and A. Yener, "The general Gaussian multiple-access and two-way wiretap channels: Achievable rates and cooperative jamming," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2735–2751, Jun. 2008.

[13] S. Bakirtas, M. R. Bloch, and E. Erkip, "Pilot-attacks can enable positive-rate covert communications of wireless hardware Trojans," *available on arXiv*, 2024.

[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 2006.

[15] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*. MIT Press, 2022.

[16] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, 2003.

[17] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1468–1489, 1999.

[18] S. V. Hanly and D. N. C. Tse, "Multiaccess fading channels. ii. delay-limited capacities," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2816–2831, 1998.

[19] H. Q. Ta and S. W. Kim, "Covert communication under channel uncertainty and noise uncertainty," in *IEEE ICC 2019*. IEEE, 2019, pp. 1–6.

[20] S. Lee, R. J. Baxley, M. A. Weitnauer, and B. Walkenhorst, "Achieving undetectable communication," *JSTSP*, vol. 9, no. 7, pp. 1195–1205, 2015.

[21] T. V. Sobers, B. A. Bash, S. Guha, D. Towsley, and D. Goeckel, "Covert communication in the presence of an uninformed jammer," *IEEE TWC*, vol. 16, no. 9, pp. 6193–6206, 2017.

[22] B. Laurent and P. Massart, "Adaptive estimation of a quadratic functional by model selection," *Annals of Statistics*, pp. 1302–1338, 2000.