Amplitude-Modulating Analog/RF Hardware Trojans in Wireless Networks: Risks and Remedies

Kiruba Sankaran Subramani, *Member, IEEE*, Noha Helal, *Student Member, IEEE*, Angelos Antonopoulos, *Member, IEEE*, Aria Nosratinia, *Fellow, IEEE*, and Yiorgos Makris, *Senior Member, IEEE*

***Abstract*— We investigate the risk posed by amplitude-modulating analog/RF hardware Trojans in wireless networks and propose a defense mechanism to mitigate the threat. First, we introduce the operating principles of amplitude-modulating analog/RF hardware Trojan circuits and we theoretically analyze their performance characteristics. Subject to channel conditions and hardware Trojan design restrictions, this analysis seeks to determine the impact of these malicious circuits on the legitimate communication and to understand the capabilities of the covert channel that they establish in practical wireless networks, by characterizing its error probability. Next, we present the implementation of two hardware Trojan examples on a Wireless Open-Access Research Platform (WARP)-based experimental setup. These examples reside in the analog and the RF circuitry of an 802.11a/g transmitter, respectively, where they manipulate the transmitted signal characteristics to leak their payload bits. Using these examples, we demonstrate (i) attack robustness, i.e., ability of the rogue receiver to successfully retrieve the leaked data, and (ii) attack inconspicuousness, i.e., ability of the hardware Trojan circuits to evade detection by existing defense methods. Lastly, we propose a defense mechanism that is capable of detecting analog/RF hardware Trojans in WiFi transceivers. The proposed defense, termed Adaptive Channel Estimation (ACE), leverages channel estimation capabilities of Orthogonal Frequency Division Multiplexing (OFDM) systems to robustly expose the Trojan activity in the presence of channel fading and device noise. Effectiveness of the ACE defense has been verified through experiments conducted in actual channel conditions, namely over-the-air and in the presence of interference.**

***Index Terms*— Hardware Trojan, covert channel, IEEE 802.11a/g, adaptive channel estimation.**

I. Introduction

HARDWARE Trojans are malicious modifications introduced by an adversary in an Integrated Circuit (IC) to interfere with its legitimate operation and/or exfiltrate sensitive information. Among the numerous hardware Trojan attacks and defenses which have been developed [1]–[7], the vast majority target digital circuits. In recent years, however, hardware Trojan attacks have also been studied in the context of wireless networks, mostly using simple wireless links [8]–[11]. Indeed, wireless networks are an attractive target for hardware Trojan attacks, since they exchange information over public channels, thereby eliminating the need for physical access to their nodes. Alternatively, attacks which leak information through side channels without necessitating hardware modifications have also started to appear in the literature [12], [13].

For reasons such as conservative design, which reduces cost and ensures high manufacturing yield in the presence of process variation, practical wireless devices do not typically operate at the boundaries of their circuit and standards specifications. Therefore, a margin exists between the operating point of these devices and the aforementioned boundaries. This margin is precisely what can be exploited by hardware Trojans to stage their attack on wireless devices. Towards mitigating the security risk introduced by this margin, in [14] we presented a preliminary study consisting of (i) a defense method, called Adaptive Channel Estimation (ACE), which monitors the wireless channel characteristics to identify inconsistencies that may be caused by a hardware Trojan operation, and (ii) a practical evaluation of this defense using a hardware Trojan that was implemented in the RF front-end of an 802.11a/g transmitter. In this work, we extend the preliminary study presented in [14] by theoretically analyzing the performance characteristics of amplitude-modulating analog/RF hardware Trojans and experimentally demonstrating the Trojan-agnostic operation and effectiveness of the proposed defense using an additional Trojan example. Specifically, compared with [14], in this paper we make the following additional contributions:

- Theoretical analysis of the Trojan impact on the legitimate communication.
- Analytical determination and experimental verification of the covert communication error probability.
- Introduction of a system-level hardware Trojan attack, which is embedded in the analog circuitry of a WiFi transmitter.
- Demonstration of the Trojan-agnostic operation of the proposed ACE-based defense, as well as its effectiveness, using the additional Trojan example.
- Contrasting of the two hardware Trojan attacks and the ACE-based defense with the state-of-the-art in hardware Trojan attacks and defenses in wireless networks.

Manuscript received October 31, 2019; revised March 2, 2020; accepted April 4, 2020. Date of publication April 27, 2020; date of current version June 16, 2020. This work was supported in part by the National Science Foundation under Grant 1514050. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Georg Sigl. *(Corresponding author: Yiorgos Makris.)*

Kiruba Sankaran Subramani, Noha Helal, Aria Nosratinia, and Yiorgos Makris are with the Department of Electrical and Computer Engineering, The University of Texas at Dallas, Richardson, TX 75080 USA (e-mail: kiruba.subramani@utdallas.edu; noha.helal@utdallas.edu; aria@utdallas.edu; yiorgos.makris@utdallas.edu).

Angelos Antonopoulos is with u-blox Athens S.A., 15125 Maroussi, Greece (e-mail: aanton@utdallas.edu).

Digital Object Identifier 10.1109/TIFS.2020.2990792

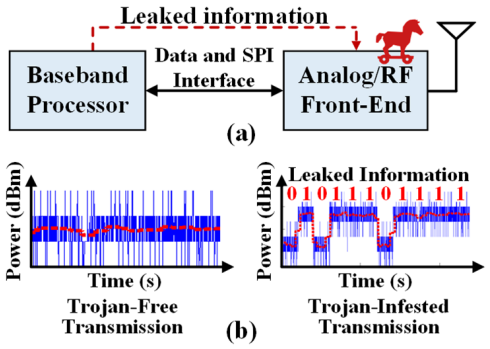


Fig. 1. Overview of amplitude-modulating hardware Trojan.

Overall, the two hardware Trojan attacks presented herein demonstrate the ability of malicious circuits to establish covert communication channels by exploiting the margins in the analog/RF front-end of a standards-compliant wireless device. While doing so, the hardware Trojan operation remains undetectable by post-fabrication tests and existing defense methods. In contrast, ACE provides the additional capability required for legitimate receivers to distinguish between a Trojan-free and a Trojan-infested communication. The proposed defense does not assume any knowledge of the hardware Trojan specifics (i.e., it is Trojan-agnostic) and its performance cannot be tampered with by an attacker, as it is implemented on the receiver side.

The rest of this paper is organized as follows. Section II presents two practical examples of amplitude-modulating analog/RF hardware Trojans and theoretically analyzes the performance of such malicious circuits in an IEEE 802.11a/g transmitter. Section III introduces the ACE-based hardware Trojan defense method. Section IV describes the experimental platform and presents experimental results evaluating the effectiveness of the two hardware Trojan attacks and the proposed defense mechanism. Section V provides a comparison to related work and conclusions are drawn in Section VI.

II. Analog/RF Hardware Trojans

The physical layer of a standards compliant wireless device consists of the baseband processor and the analog/RF front-end, as shown in Figure 1(a). The baseband processor is responsible for implementing the wireless protocol and its associated signal processing blocks. In a transmitter design, this includes performing operations such as encoding, interleaving, modulation, inverse Fourier transform and cyclic prefix insertion to convert the raw user data into a form suitable for transmission. The analog/RF front-end, on the other hand, plays a critical role in enabling the transmission and reception of the transformed user data through the designated wireless communication channel. The front-end includes all the circuitry between the baseband processor and the antenna, consisting of data converters, filters, baseband amplifiers, mixers and power amplifiers.

In the past, analog/RF hardware Trojan attacks have predominantly been demonstrated using standalone circuits [15]–[18]. Among the handful of research works that focused on wireless devices, the majority use simple links rather than standards-compliant devices to demonstrate the threat. Beyond the vulnerabilities of these simple links, practical wireless devices have inherent margins in their hardware, which can facilitate malicious hardware Trojan attacks. These margins exist between the operating point of the device and the boundaries defined by its circuit and standards specifications. There are several reasons for the existence of these margins:

- Circuits are designed and rated conservatively to ensure high yield in the presence of process variation.
- Transmitter / receiver separation distance is not always at the edge of the respective range for a power setting, producing a gap between the required and the actual transmission power.
- Channel conditions are dynamic and are often imperfectly known to the transmitter and receiver.

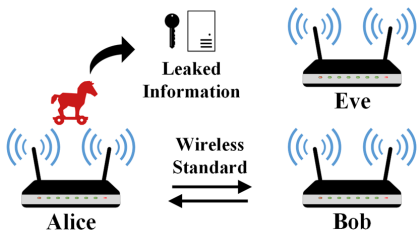


Fig. 2. Threat model.

In the analog/RF front-end, in particular, hardware Trojans can exploit these margins to systematically modify the parameters of the transmitted signal, such as the amplitude, frequency, phase or combinations thereof, to leak sensitive information from the targeted device. Figure 1(b) shows one such example comparing a Trojan-free and a Trojan-infested transmission, where the malicious circuit has embedded the leaked information in the transmitted signal amplitude, while remaining compliant with the wireless protocol and the design specifications. Such variations in the amplitude can be created by a hardware Trojan that has been realized through some form of circuit- or system-level modification in one or more blocks of the transmitter chain. Meanwhile, an adversary who is privy to how the leaked data is embedded in the transmitted signal amplitude, can retrieve it through a rogue receiver which observes the systematic variations in the received signal power.

To establish a reliable covert communication based on amplitude-modulation, while simultaneously minimizing the Trojan impact on the legitimate communication, the hardware Trojan performance parameters need to be chosen carefully. To this end, in this section we first introduce the threat model and present two practical instances of amplitude-modulating analog/RF hardware Trojan circuits in the context of a standards-compliant wireless network. Next, we theoretically analyze the performance of such Trojan circuits when embedded in a wireless device and we discuss the trade-offs involved in the design of such malicious circuits.

A. Threat Model

The threat scenario considered in this work is shown in Figure 2. Here, Alice and Bob are two standards-compliant wireless devices that have a legitimate communication

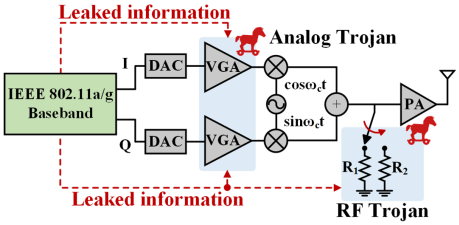


Fig. 3. Simplified model of the Analog/RF hardware Trojans.

established between them. Unbeknownst to Alice, her wireless hardware has been tampered with by an attacker, who has introduced a hardware Trojan circuit. The centerpiece of the malicious entity resides in the analog/RF front-end of her device, where it systematically modifies the transmission power to exfiltrate secret information. Meanwhile, Eve, the third wireless device in the example which represents the rogue receiver, observes the systematic distortions and retrieves the leaked data.

The threat model assumes that the hardware Trojan has been implanted either during design or during IC fabrication of the transmitter, and may be used once the device is deployed. The data leaked through the covert channel, which can be an encryption key, plaintext, or any other sensitive information, resides in the baseband part of the wireless device and is forwarded to the analog/RF front-end –where the attack is staged– by additional malicious modifications, as shown in Figure 1. Thereby, the leaked information bits are embedded in the transmitted signal through subtle amplitude modifications.

B. Hardware Trojan Examples

We now describe two examples of amplitude-modulating analog/RF hardware Trojans in an IEEE 802.11a/g network. In [14], we introduced a hardware Trojan that was implemented in the RF front-end circuitry of a wireless transmitter, where the malicious circuit exploits the process variation margins of the targeted IC to stage its attack. In this work, we extend this contribution by also presenting a second hardware Trojan example that is realized in the analog domain, where the malicious circuit manipulates the programmable gain stages of a wireless device. The two Trojan attacks, shown in Figure 3, leak sensitive information to a rogue receiver through imperceptible variations in the transmitted signal power. The Trojan operation incurs negligible impact on the legitimate communication and does not violate any wireless standard specifications, as demonstrated in Section IV.

1) RF Trojan: In a wireless transceiver, input/output ports of RF ICs are terminated in a load impedance that is matched to the impedance of adjacent stages to avoid signal reflection. The value of this impedance is typically 50 Ω.[1] In practical devices, however, due to parasitics and imperfections associated with the manufacturing process, it is not possible to achieve a perfect impedance match between successive blocks in the chain. As a result, a tolerance level in the form of return loss is provided in the IC specifications to account for this mismatch. Such margin, however, also provides room for attackers to manipulate the termination impedance of the targeted IC to leak sensitive information. The RF Trojan, proposed in [14] and shown in Figure 3, uses this principle to stage its attack in a wireless transmitter. Essentially, the Trojan circuit uses a Single Pole Double Throw switch and a pair of slightly different resistors to systematically alter the input termination impedance of the power amplifier based on the bit values of the leaked data. This operation creates subtle variations in the power amplifier's input reflection coefficient and, thereby, the transmitted signal power. Details of this malicious circuit, which we implemented on a Printed Circuit Board (PCB), are provided in Section IV.

2) Analog Trojan: Wireless devices use multiple gain stages in the transmitter chain to satisfy linearity and to achieve the desired performance specifications. These stages are, often, programmable and are connected to the baseband processor through a Serial Peripheral Interface to facilitate boot-up configuration. Therefore, similar to the RF Trojan attack described above, information leaked from the baseband can be used to systematically modify the transmission power by exploiting the programmable gain stages. The second Trojan example shown in Figure 3 is based on this principle, where the malicious entity systematically changes the gain of the Variable Gain Amplifiers (VGAs) to create minute variations in the transmission power in accordance with the leaked bits. A detailed description of this Trojan circuit is provided in Section IV, along with its implementation details in an IEEE 802.11a/g transmitter.

3) Rogue Receiver: The rogue receiver knows that the covert channel is established by modulating transmission amplitude. Based on this knowledge, it leverages the Received Signal Strength Indicator of the WiFi receiver architecture to extract the leaked information bits. Specifically, the rogue receiver continuously tracks the received signal power over the duration of one leaked information bit. Thereby, subtle discrepancies in the power profile are identified, from which the leaked information bits are retrieved based on a threshold value. Details of the rogue receiver, which we implemented for demonstrating successful retrieval of the leaked data, are provided in Section IV.

C. Theoretical Analysis of Hardware Trojan Impact

We now present a theoretical analysis of the performance characteristics of amplitude-modulating analog/RF hardware Trojans in wireless communication systems, independent of the Trojan implementation. This study has two aspects: understanding the impact of the hardware Trojan operation on the legitimate communication, and characterizing the error probability of the Trojan-induced covert channel.

1) At the Legitimate Receiver: The packet error rate (PER) of a convolutional code followed by an M-ary Quadrature Amplitude Modulation (QAM) modulator in an Additive White Gaussian Noise (AWGN) channel is bounded by [19]:

$$\text{PER} \leq (L - d_{free}) \sum_{d=d_{free}}^{\infty} A(d) P_2(d) \qquad (1)$$

[1]This value is used only as a reference; the hardware Trojan attack *principles* described herein are independent of this value.

TABLE I

DISTANCE SPECTRUM OF THE CONVOLUTIONAL CODE WITH RATE 1/2 USED IN IEEE 802.11A/G

d	10	12	14	16	18
$A(d)$	11	38	193	1331	7275

where L is the packet length, d_{free} is the minimum free distance of the convolutional code, the distance spectrum $A(d)$ is the number of valid codewords that are within a distance d from the all-zero codeword and $P_2(d)$ is the probability that an incorrect codeword is selected at the receiver with distance d from the correct codeword. For different M-ary QAM modulation, $P_2(d)$ is given by:

$$P_2(d)_{\text{BPSK}} = Q\left(\sqrt{2dR_c\frac{E_b}{N_0}}\right) \tag{2}$$

$$P_2(d)_{\text{QPSK}} = Q\left(\sqrt{2dR_c\frac{E_b}{N_0}}\right) \tag{3}$$

$$Q\left(\sqrt{\frac{36}{5}dR_c\frac{E_b}{N_0}}\right) \leq P_2(d)_{16\text{QAM}} \leq Q\left(\sqrt{\frac{4}{5}dR_c\frac{E_b}{N_0}}\right) \tag{4}$$

where Q$(\cdot)$ is the standard normal Q function and $\frac{E_b}{N_0}$ is the Signal-to-Noise Ratio (SNR) per bit. We denote the PER in (1) by $\text{PER}_{(E_b)}$.

In this threat model, the Trojan inserts rogue data into the transmitted signal through subtle modification of its power characteristics. Without loss of generality, we can assume that the rogue bits are equally probable to be "0" or "1". Therefore, the transmitted symbol has an average bit energy E_{b1} for a "1" rogue bit or E_{b0} for a "0" rogue bit with probability 1/2. Assuming each rogue message bit can be zero or one with equal probability, the packet error rate at the legitimate receiver in the presence of the Trojan is:

$$\text{PER}_{\text{BOB}} = \frac{1}{2}\text{PER}_{(E_{b1})} + \frac{1}{2}\text{PER}_{(E_{b0})} \tag{5}$$

For demonstration purposes, we use a rate 1/2 convolutional code adopted in the IEEE 802.11a/g standard [20]. This code is characterized by a constraint length of 6, generator polynomials $g_0 = 1011011$ and $g_1 = 1111001$ and distance spectrum as shown in Table I. We point out that the summation in (1) is dominated by the first few terms, therefore, the five values presented in Table I are sufficient for PER calculation.

Figure 4 shows the theoretical results characterizing the hardware Trojan impact on the legitimate communication. In the figure, the PER of a binary convolutional code in an AWGN channel is shown for the clean and Trojan-infested transmitter under Binary Phase Shift Keying (BPSK), Quadrature Phase Shift Keying (QPSK) and 16-QAM modulations as a function of the Singal-to-Noise Ratio (SNR) observed at the legitimate receiver. For a Trojan signal amplitude $\Delta = E_{b1} - E_{b0} = 1\,\text{dB}$ between the Trojan levels, the plot reveals that a Trojan-infested transmitter will require approximately 0.3 dB more power to achieve an error probability of 10^{-3} compared with the clean transmitter. This is

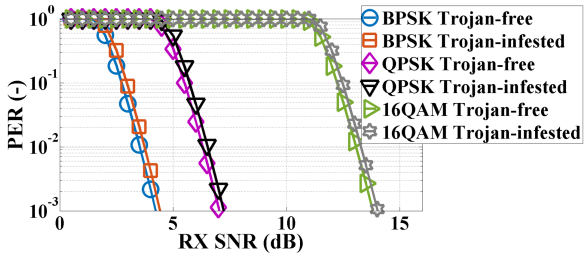


Fig. 4. Theoretical results for clean and contaminated transmitter.

well within the variations introduced by the wireless channel and device noise. Therefore, this Trojan cannot be easily distinguished from ubiquitous channel imperfections. Moreover, the observed Trojan impact is consistent across modulation schemes, thereby verifying the robustness of the hardware Trojan attack.

2) At the Rogue Receiver: Consider a received Amplitude-Shift Keying (ASK) waveform, consisting of a signal component and a noise component

$$y(t) = m(t)\cos(w_0 t) + n(t) \tag{6}$$

where $m(t)$ is the amplitude of the carrier wave and $n(t)$ is the noise amplitude. At the receiver, this signal goes through a matched filter and is sampled. For simplicity, in the following we only refer to sampled values. The sampled *received* signal is denoted by r, and the sampled *transmitted* signal by m, which can take values $\{V_0, V_1\}$. We assume $V_1 > V_0$. The underlying bits represented by these amplitudes are $b \in \{0, 1\}$. The relation of transmitted and received sampled values is:

$$r = \sqrt{(m + n_1)^2 + n_2^2}$$

where n_1, n_2 are Gaussian random variables with mean zero and variance N_0.

If the value of r is above a certain threshold (to be optimally determined), then the receiver decides a "1" was transmitted; otherwise, it decides a "0" was transmitted. It can be shown that the PDF of r is given by [21]:

$$f_R(r|m = V_1) = \frac{r}{N_0}e^{\frac{-(r^2+V_1^2)}{2N_0}} I_0\left(\frac{rV_1}{N_0}\right) \tag{7}$$

$$f_R(r|m = V_0) = \frac{r}{N_0}e^{\frac{-(r^2+V_0^2)}{2N_0}} I_0\left(\frac{rV_0}{N_0}\right) \tag{8}$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind of zeroth order. Assuming that "0" and "1" are equiprobable, the optimum detection according to the maximum likelihood is identical to Maximum a Posteriori estimation (MAP):

$$\frac{f_R(r|m = V_1)}{f_R(r|m = V_0)} \underset{\text{b=0}}{\overset{\text{b=1}}{\gtrless}} 1 \tag{9}$$

The left hand side (likelihood ratio) for our problem is a monotonic function of r. Therefore the solution of

$$f_R(r|m = V_1) = f_R(r|m = V_0)$$

namely $\eta \triangleq \frac{V_0+V_1}{2}$, yields the detection rule:

$$\hat{b} = \begin{cases} 0 & \text{if } r < \eta \\ 1 & \text{if } r > \eta \end{cases} \tag{10}$$

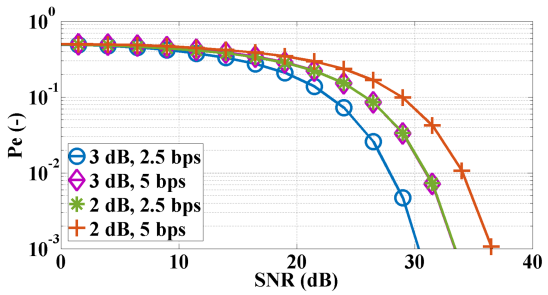


Fig. 5. Theoretical results characterizing the reliability of the Trojan communication.

The overall error probability is given by integrating the conditional PDFs over the error regions:

$$\begin{aligned} P_e &= \frac{1}{2}\int_0^{\eta} f_R(r|m=V_1)dr + \frac{1}{2}\int_{\eta}^{\infty} f_R(r|m=V_0)dr \\ &= \frac{1}{2}\left[1 - Q_1\left(\frac{V_0}{\sqrt{N_0}}, \frac{\eta}{\sqrt{N_0}}\right) + Q_1\left(\frac{V_1}{\sqrt{N_0}}, \frac{\eta}{\sqrt{N_0}}\right)\right] \\ &= \frac{1}{2}\left[1 - Q_1\left(\sqrt{\frac{E_{b0}R_b}{N_0}}, \frac{\eta}{\sqrt{N_0}}\right) + Q_1\left(\sqrt{\frac{E_{b1}R_b}{N_0}}, \frac{\eta}{\sqrt{N_0}}\right)\right] \end{aligned} \tag{11}$$

where R_b is the rogue data rate and $Q_1(\cdot,\cdot)$ is the Marcum Q-function

$$Q_1(a,c) = \int_c^{\infty} x e^{-\frac{x^2+a^2}{2}} I_0(ax)dx \tag{12}$$

The sufficient statistic (equivalent SNR) for the Trojan channel is $\frac{V_1-V_0}{\sqrt{N_0}}$. During experiments, however, controlling the ratio V_1/V_0 is much simpler and more precise compared with controlling the difference $V_1 - V_0$. Therefore, to facilitate comparison of theory and experiment, we report error probabilities while stepping through different values of V_1/V_0 (expressed in dB). Intuitively, the separation of the two levels can be grasped via either their difference or their ratio. Furthermore, it is straightforward to verify that, considered together with the SNR of the legitimate channel, controlling one is equivalent to controlling the other.

Figure 5 plots the error probability of Trojan communication from Eq. (11). In this figure, we explore the tradeoff between Trojan transmission rate, Trojan probability of error, and amplitude of Trojan infestation. The horizontal axis represents the total received SNR, which is the total signal power (averaged over Trojan amplitude manipulations) divided by noise power.

The Trojan communication rate is either 2.5 bps or 5 bps. It is observed that higher Trojan rates will either lead to more Trojan errors, or require a larger amplitude modulation that would increase the probability of the Trojan being detected. Further, the analysis indicates that reliability of covert communication is improved as rogue data rates are reduced and as Trojan amplitude is increased, but the latter comes at the cost of diminished Trojan inconspicuousness, as mentioned earlier.

III. Defense Mechanism

The amplitude-modulating hardware Trojan examples and their theoretical analysis described in the previous section provide valuable insight regarding the threat posed by malicious hardware in wireless networks. Ultimately, however, our objective is to improve the overall security of wireless networks by developing appropriate countermeasures. To this end, in this section we propose a defense mechanism that is capable of detecting analog/RF hardware Trojans in WiFi transceivers.

Wireless channels in a wideband communication network are frequency-selective and time-varying. Therefore, WiFi transceivers use *channel estimation algorithms* for coherent detection and decoding of the data packets. These algorithms use training sequences in the packet preamble, along with the pilot symbols, to periodically estimate the channel conditions. In current receivers, however, the channel estimation algorithms bundle together any malicious disturbances introduced by a hardware Trojan with the inherent non-idealities of the wireless channel. The proposed ACE defense overcomes this limitation by exploiting the slow-fading characteristic of an indoor communication channel to distinguish between the channel-induced and the Trojan-induced impact on the estimated coefficients. This allows the defense method to detect any Trojan operation in a wireless network, regardless of the attack specifics.

A. Channel Estimation

Performance of a wireless network is affected by a variety of channel impairments such as noise, path loss, fading and multi-user interference. These factors not only affect the reliability and data rate of the network, but also create uncertainties in the communication, which leave margins for an adversary to stage an attack. Meanwhile, OFDM is commonly used in modern wireless networks due to its high spectral efficiency, throughput and ability to address the problems of inter-symbol and inter-carrier interference caused by multi-path propagation. Here, a large number of closely spaced orthogonal sub-carriers are used to transmit the user-data. This produces a multi-carrier spectrum in the frequency-domain, an inverse Fast Fourier Transform (FFT) of which produces the OFDM symbols in the time domain. Further, OFDM inserts pilot sub-carriers and preamble within the transmitted data to facilitate channel estimation at the receiver, as shown on the left side of Figure 6. Based on the location of the inserted pilot tones, OFDM supports a block-type or comb-type channel estimation. In a block-type channel estimation, the pilot tones are inserted in all the sub-carrier locations of an OFDM symbol and the channel coefficients are estimated once for every M symbols. Hence, a block-type channel estimation is typically used under slow-fading channel conditions. On the other hand, in a comb-type channel estimation, the pilot tones are inserted into each transmitted symbol at uniform locations. This allows the system to estimate the channel coefficients on a per symbol basis, which is essential when operating under fast fading channel conditions. In either type, the fading channel of an OFDM system can be viewed as a 2D lattice in the time-frequency plane, as shown on the left side of Figure 6, where the channel characteristics are estimated at the pilot positions using least squares (LS), minimum mean-square

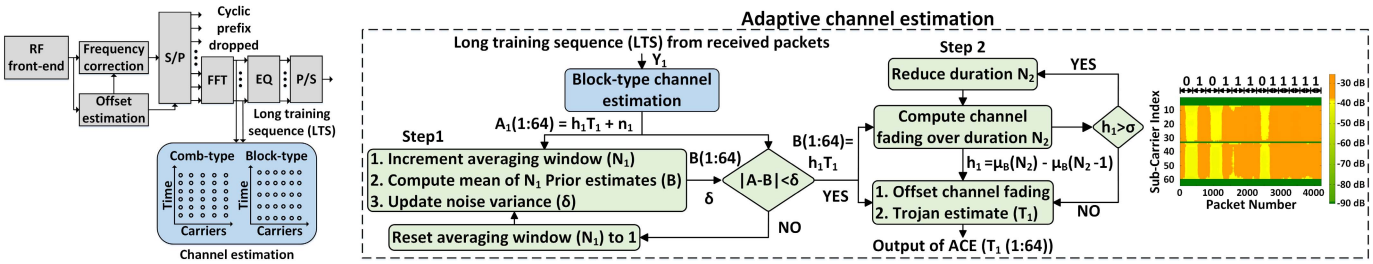


Fig. 6. ACE-based hardware Trojan detection method.

error (MMSE), or modified MMSE [22] and, subsequently, interpolated between the pilots using linear, second-order or low-pass interpolation techniques.

Given the slow-fading nature of WiFi, especially in indoor environments [23], we implement a block-type channel estimation algorithm wherein the channel characteristics vary at a much slower rate compared with the packet/symbol duration. The objective of this algorithm, which is shown in Equation (13), is to estimate the channel matrix (**h**), given the matrix (**x**) containing the packet preamble and pilot symbols - which is known to the receiver - and the received signal (**y**), in the presence of additive white Gaussian noise (**n**):

$$y = hx + n \tag{13}$$

B. Adaptive Channel Estimation

In a Trojan-free communication, the transmitted signal (**x**) is predominantly affected by the channel matrix (**h**), which is a multiplicative term, as shown in Equation (13). However, in a Trojan-infested communication, the hardware Trojan alters the transmitted signal such that **x** is scaled by an additional term **T**, which represents the impact of the hardware Trojan, as described by Equation (14).

$$y = hTx + n \tag{14}$$

Since the channel matrix (**h**) and the hardware Trojan impact (**T**) are two multiplicative terms that affect the transmitted signal, an unsuspecting legitimate receiver treats the two unknown factors as a single entity, estimates its corresponding coefficients and recovers the received signal. The additional capability required for the legitimate receiver to distinguish the Trojan activity from channel conditions and Gaussian noise is provided by the proposed defense mechanism, which is shown in Figure 6 and explained next.

The proposed defense, namely ACE, receives long training sequences (LTS) from the packet preamble, which are processed in two steps to identify the Trojan activity (**T**). In the first step, ACE removes the additive Gaussian noise (**n**) using adaptive averaging. Since the legitimate receiver is not privy to the Trojan implementation details and its throughput, the defense algorithm computes the noise variance (δ) of the incoming channel estimates and uses it as a metric to adjust the averaging window duration (N_1). In other words, ACE computes the difference between the incoming channel estimates (A) and the mean of the past channel estimates (B). Thereby, if the difference is smaller than δ, the averaging window size is increased. In contrast, if the difference is larger than δ, ACE considers it an anomaly and resets the averaging window size. As a result of the averaging, the additive Gaussian noise (**n**) is removed from the received signal at the end of the first step.

In the second step, ACE separates the Trojan activity (**T**) and the channel matrix (**h**). To determine **h**, the defense computes the mean of the de-noised channel estimates over a duration (N_2), which is chosen large enough ($\sim$3 ms) with respect to the slow-fading nature of indoor communication channels. It is important to note that, due to slow fading, $N_2 \gg N_1$. The computed mean value, namely μ_B in Figure 6, is compared between successive intervals based on a threshold (σ), whose value is obtained from the slow fading indoor channel model [23]. This enables the defense to determine the rate at which the channel parameters vary and, accordingly, adjust N_2 to enhance the estimation of **h**. Once **h** is determined, it is removed from the de-noised channel estimates from Step 1 to reveal the Trojan activity (**T**).

An example of the ACE output is shown on the right-hand side of Figure 6. Here, the x-axis represents the received packet number, which is in the time-domain, and the y-axis denotes the sub-carrier index, which is in the frequency domain. For each received packet, the defense algorithm produces an output that spans across the 64 sub-carrier locations, which is plotted as a heat-map, as shown in the figure. In the sample output, the heat-map has a dark-green color at sub-carrier locations 1:6, 32 and 60:64. These regions correspond to the guard-band and the DC sub-carrier locations of the data packet, where the signal power is extremely small (-90 dBm). For the remaining 52 sub-carriers (data and pilots), the heat-map exhibits different colors according to their magnitude levels.

Detection of the Trojan activity is, subsequently, based on the amount of variance observed in the generated heat-map. Specifically, in a Trojan-free communication, where there is no suspicious activity, removing the channel fading and additive Gaussian noise from the channel estimates results in a uniformly colored heat-map. In contrast, in a Trojan-infested communication, where the hardware Trojan systematically manipulates the transmitted signal to leak information, the heat-map will have distinct colors between successive leaked bits, thereby revealing the Trojan presence.

For the final detection decision, the elements of the vector **T**, corresponding to WiFi subcarriers, are aggregated into one decision variable via averaging or weighted averaging, over a moving time window. Then, they are compared with a threshold whose value depends on the trade-off between false

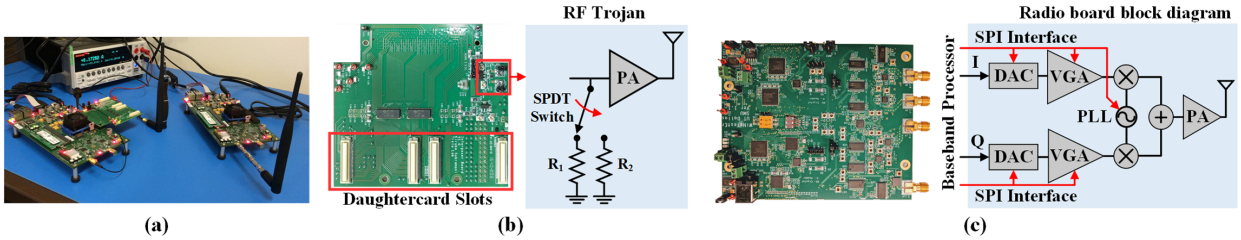


Fig. 7. Experimental platform: (a) WARP-based setup, (b) Interposer board and (c) RF board.

alarm and missed detection rates that is appropriate for the individual application or circuit.

IV. Experimental Results

To assess the performance of the two hardware Trojan attacks and the proposed defense method, we put together an experimental platform based on WARP v3 boards. Figure 7(a) shows the platform where one board functions as the transmitter and the other as the receiver, operating under the IEEE 802.11a/g protocol. In addition, the setup uses custom hardware, namely an interposer and a radio board, on which the two hardware Trojan examples have been realized. During our experiments, the custom boards are mounted on the transmitter node using the dedicated daughtercard slot.

Interposer: The first board, shown in Figure 7(b), houses the RF Trojan that was introduced in Section II-B. This board has a Trojan-free and a Trojan-infested version of the power amplifier to facilitate performance characterization of the communication under these two scenarios. In both cases, the input of the power amplifier is connected to the WARP RF front-end output using a coaxial cable. The RF Trojan, which is realized on this board, uses a Single Pole Double Throw switch and a pair of slightly different termination resistors, as shown in Figure 7(b). The leaked information, which controls the switch of the Trojan circuit, comes from the baseband processor that is implemented on the FPGA. During Trojan operation, the leaked bit value determines which of these two resistors is connected in parallel to the input of the power amplifier. This act slightly modifies the input impedance of the power amplifier and, thereby, creates an imperceptible variation in the transmitted signal power. In this example, resistor values of "R_1" $= 0.8\,\text{k}\Omega$ and "R_2" $= 30\,\text{k}\Omega$ were used, corresponding to leaked information bits "0" and "1", respectively. When terminated by "R_2", the input impedance of the power amplifier and the transmitted signal power are close to an ideal $50\,\Omega$ condition. Whereas, when terminated by "R_1", the input impedance changes to $47.5\,\Omega$, resulting in a 5% reduction in the transmitted power. This difference can then be exploited by an attacker to establish a Trojan communication to leak information in an unauthorized fashion.

When integrated in a WiFi transceiver, the area and power overhead of the hardware Trojan becomes negligible compared with the power amplifier circuit. To verify this, we designed a Class A power amplifier and the hardware Trojan circuit in Cadence using GlobalFoundries' 130nm RF CMOS process. In this design, the Trojan circuit occupies an area of $9.56\,\mu\text{m}^2$ and consumes power of 0.72 nW, which is negligible when compared with an area of $0.57\,\text{mm}^2$ and a power of 415.8 mW of the power amplifier circuit. This footprint becomes even smaller when compared with the overall transceiver area and power consumption.

Radio Board The second board, shown in Figure 7(c), houses the analog Trojan that was introduced in Section II-B. The radio board implements a 0-6 GHz RF transceiver, consisting of data converters, VGAs, mixers, filters and a Phase Locked Loop. During normal operation, the radio board is interfaced to the WARP board with the help of the interposer. This allows the radio board to serve as an RF front-end to the WARP board, whose baseband input/output connections and Serial Peripheral Interfaces are tied to the baseband logic running on the FPGA through the dedicated daughtercard slots. These system-level connections are, however, susceptible to modifications through tampering of the hardware or through exploitation of vulnerabilities in the device firmware. The second hardware Trojan example introduced in Section II-B is realized based on this principle. Here, the Trojan circuit resides in the MicroBlaze processor of the FPGA and exploits the Serial Peripheral Interface to the VGAs located in the In-phase (I) and the Quadrature-phase (Q) signal paths to systematically modify their gain and, thereby, leak secret information. Specifically, when leaking a "1" bit, the gain of the VGAs remains unchanged, whereas when leaking a "0" bit, the gain is reduced to create a minute variation in the transmitted power.[2] Since this analog Trojan leverages existing system level connections along with minor modifications to the FPGA code, the overall area and power overhead that it incurs is negligible.

In the following subsections, we experimentally assess the performance of the two hardware Trojan attacks and the proposed defense method using this setup.

A. Attack Effectiveness

Our first experiment demonstrates the ability of hardware Trojans to successfully leak information bits to an adversary through a covert channel. An adversary, who is aware of the Trojan implementation details (i.e., amplitude modulation mechanism and throughput) can use a rogue receiver to retrieve the leaked information bits. For this experiment we used a second WARP board to implement a rogue receiver which averages the received signal power and thresholds it to retrieve the leaked bits, as explained in Section II-B.

[2]The gain of both VGAs needs to be reduced identically to prevent amplitude imbalance between the I and Q paths.

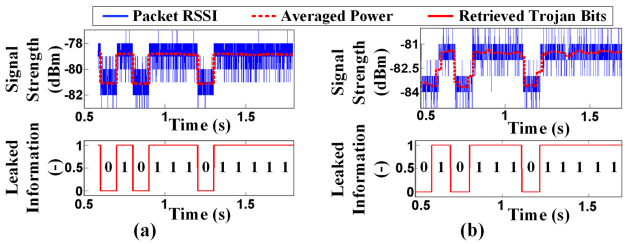


Fig. 8. Decodability of the leaked information: (a) RF Trojan and (b) Analog Trojan.

1) Effectiveness of RF Trojan: Figure 8(a) shows the signal strength of the received contaminated signal, i.e., the signal conveying both the legitimate and the leaked information, from an RF Trojan-infested transmitter. The plot is shown for a duration of 2 s, during which the RF Trojan leaks information bits "0101 1101 1111" by systematically modifying the transmitted power. An adversary recovers the leaked information by averaging the received signal strength over the duration of one leaked information bit and attributes an increase / decrease in the average value to a leaked bit "0" / "1" based on a threshold, which is currently set at 1.5 dB. Figure 8(a) demonstrates this Trojan activity, where the 24 bits leaked by the malicious entity are successfully retrieved by the rogue receiver.

2) Effectiveness of Analog Trojan: The above experiment was repeated using the analog Trojan, where the leaked information bits were, again, set as "0101 1101 1111". Figure 8(b) shows the received contaminated signal for a duration of 2 s, where the analog Trojan has introduced systematic variations in the signal power by changing the gain of the VGAs in the I and Q path. Similar to the RF Trojan example, an adversary recovers the leaked information from the received contaminated signal by using the rogue receiver, where an increase/decrease in the average power is attributed to a leaked information bit "0"/"1" based on a threshold value, which is again set at 1.5 dB. Effectiveness of the analog Trojan is shown in Figure 8(b), where all the 24 leaked information bits are successfully retrieved by the rogue receiver.

B. Reliability of Covert Communication

Reliability of the covert communication was evaluated using uncoded leaked bits and under the assumption of an AWGN channel. For this experiment, data rates of 2.5 bps and 5 bps were considered for the leaked information and the Trojan amplitude was varied from 2 dB to 3 dB. Accordingly, the bit error rate (BER) of the covert communication was analyzed as a function of the SNR observed at the rogue receiver and the corresponding results are shown in Figure 9.[3] When analyzing the impact of the rogue data rate on the reliability of the covert communication, we observe that the plot corresponding to a data rate of 2.5 bps experiences the lowest BER. Since the hardware Trojan leaks information by modifying the signal strength of the transmitted data packets, increasing the rogue data rate reduces the number of signal strength values representing each leaked bit by the same proportion. Therefore, at a fixed Trojan signal amplitude, increasing the data rate of the covert communication results in a higher BER. The results presented in Figure 9 illustrate this effect. Specifically, when comparing the plots corresponding to the 2.5 bps and the 5 bps rogue data rates, respectively, for a 2 dB Trojan signal amplitude, we observe that the latter requires an additional signal power of 3 dB to achieve a BER of 10^{-3}.

[3]While this experiment evaluates covert channel performance when the legitimate device uses BPSK, similar results were observed for other modulation schemes as well. This is because the rogue receiver only relies on discrepancies in the signal power to extract the leaked information.

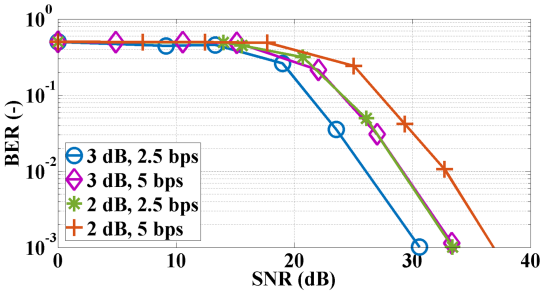


Fig. 9. Reliability of the covert communication - Impact of rogue data rate and Trojan signal amplitude.

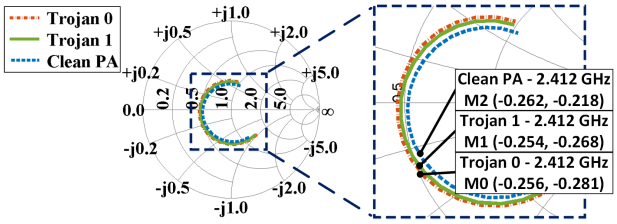


Fig. 10. Impact of the RF Trojan on S_{11}.

The effect of Trojan signal amplitude on the reliability of the covert communication is also presented in Figure 9. For this experiment, the data rate of the Trojan communication was fixed at 5 bps and the Trojan signal amplitude was varied from 2 dB to 3 dB.

C. Impact on Legitimate Communication

While an adversary is able to successfully retrieve the leaked information, operation of the hardware Trojan incurs negligible impact on the legitimate communication. To corroborate this, in this section we demonstrate that the transmitter performance characteristics and the communication error probability remain practically unaffected by the Trojan operation.

1) Impact of RF Trojan: Since the RF Trojan manipulates the input termination impedance to modulate the transmission power, in Figure 10 we plot the input reflection coefficient S_{11} of the Trojan-free and Trojan-infested power amplifier on a Smith chart. In the figure, the plots corresponding to the hardware Trojan display a slight shift when compared with the Trojan-free scenario. Such minute shift is well-within the process variation margins and, therefore, it cannot be uniquely attributed to the presence of a hardware Trojan.

Similarly, the impact of the RF Trojan on the legitimate communication is characterized in Figure 11(a). Here, the error probability, expressed in terms of PER, is analyzed with respect to the receiver's SNR. In this analysis, Trojan-free

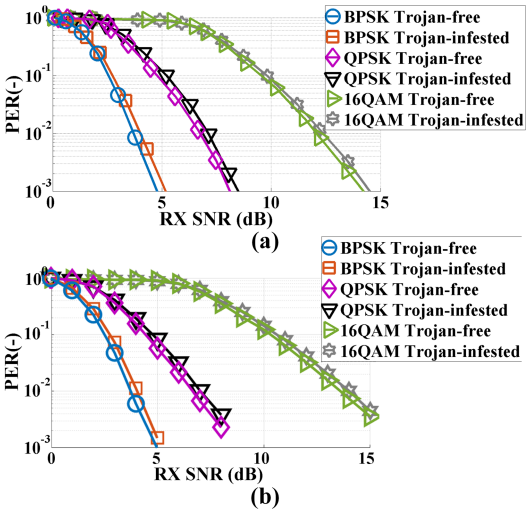


Fig. 11. Trojan impact on legitimate communication: (a) RF Trojan and (b) Analog Trojan.

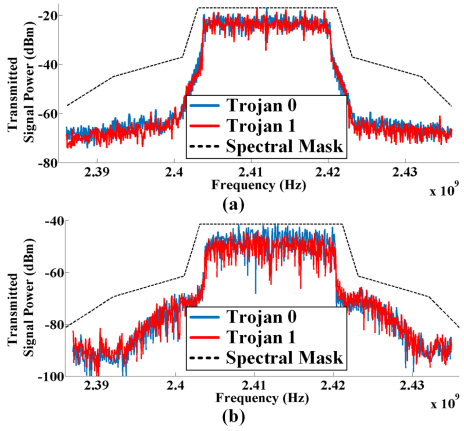


Fig. 12. Measured WiFi spectrum for: (a) RF Trojan and (b) Analog Trojan.

and Trojan-infested communications operating under BPSK, QPSK and 16-QAM have been considered, all with a coding rate of 1/2. As shown in Figure 11(a), the Trojan-infested communication requires a slight increase in SNR to achieve a target PER, as compared with the Trojan-free communication. For example, under BPSK modulation, the Trojan-infested communication requires a 0.4 dB additional signal power to achieve a PER value of 10^{-2}. Given the many uncertainties of wireless communication, such slight increase cannot be uniquely attributed to the presence of a hardware Trojan. A similar PER trend is observed for QPSK and 16-QAM in Figure 11(a), thereby verifying inconspicuousness of the RF hardware Trojan across modulation schemes.

2) Impact of Analog Trojan: Since the analog Trojan leverages the Serial Peripheral Interface to modify the VGAs' gain to leak information, it does not physically alter the transmitter hardware. Moreover, the attack is staged at the baseband frequency where S-parameters are not a part of the VGA specifications; hence, the Smith chart analysis does not apply to this attack. Therefore, the impact of the Analog Trojan on the legitimate communication has only been evaluated through the PER vs. SNR plot of Figure 11(b). The experiment considers the Trojan-free and Trojan-infested communications under the three modulation schemes mentioned earlier. Similar to the RF Trojan attack, operation of the Analog Trojan incurs a 0.3 dB increase in SNR to achieve the target PER of 10^{-2}. Once again, a consistent PER trend is observed across modulation schemes, thereby verifying inconspicuousness of the attack.

To further support our Trojan inconspicuousness claims, we note that, despite hardware imperfections and measurement noise, the experimental results of Figure 11 are remarkably consistent with the theoretical analysis results of Figure 4.

D. Detection Evasion

In this section, we evaluate existing defense methods in detecting the two hardware Trojan attacks presented herein. Post-production tests are typically the first line of defense when it comes to hardware Trojan detection. In case of WiFi transceivers, Spectral Mask and Error Vector Magnitude (EVM) are two specification tests that are performed once a device is fabricated. Similarly, Statistical Side-Channel Fingerprinting (SSCF) is one of the most popular defense methods available in literature for hardware Trojan detection. In the following subsections, we examine effectiveness of these methods in detecting the two hardware Trojan examples.

1) Spectral Mask: Figure 12(a) shows the transmitted signal spectrum from an RF Trojan-infested transmitter that was measured using a Tektronics MDO4104-6 spectrum analyzer. In the figure, the two signal spectra corresponding to the Trojan operation states (i.e., when leaking a "0" and when leaking a "1") are centered at 2.412 GHz and occupy a signal bandwidth of 20 MHz. The two signal spectra are well within the margins specified by the IEEE 802.11a/g standard [20], thereby indicating that the Trojan operation does not introduce any non-linearity into the transmitted signal. The same experiment was performed for the Analog Trojan and the results are shown in Figure 12(b), where the Trojan operation is once again fully compliant with the wireless standard.

2) Error Vector Magnitude: Our next experiment analyzed the EVM of the RF Trojan-infested communication with respect to the transmission power for BPSK, QPSK and 16-QAM. Figure 13 shows the corresponding results, along with the IEEE 802.11a/g specifications. As shown by this analysis, the hardware Trojan evades detection, since the measured EVM values are well within the margins specified by the 802.11a/g standard. A similar observation can be made for the analog Trojan, whose results are also shown in Figure 13.

In short, the experimental results corresponding to spectral mask and EVM, reveal the inability of such post-production tests to detect the presence of the hardware Trojans.

3) Statistical Side-Channel Fingerprinting: One of the most successful methods for hardware Trojan detection is SSCF [8], [11], [24]–[26]. In this method, a one-class classifier is trained to distinguish between Trojan-free and Trojan-infested circuits based on parametric measurements. To evaluate its effectiveness in detecting the RF Trojan, we generated a synthetic population of 1000 Trojan-free and 500 Trojan-infested

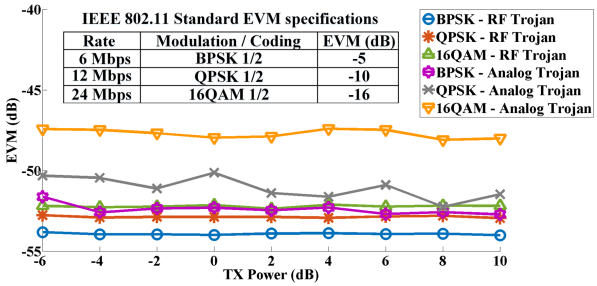


Fig. 13. Error vector magnitude test.

TABLE II
SVM WITH RBF KERNEL AGAINST RF TROJAN: (A) 6D INPUT DATASET AND (B) 12D INPUT DATASET

(a)

Actual \ Predicted	Clean	Infested
Clean	0.501	0.499
Infested	0.604	0.396

(b)

Actual \ Predicted	Clean	Infested
Clean	0.516	0.484
Infested	0.645	0.344

transmitters, using Monte Carlo simulation of the Class-A power amplifier which we designed in GlobalFoundries' 130 nm process, and embedding each of those in the block-level RF transmitter implemented in Matlab Simulink. For each device in our synthetic population, we performed an identical transmission at six different power levels between −15 dBm and +15 dBm and we measured the total transmitted power. We then used the measurements from a randomly selected 50% of the Trojan-free population (i.e., 500 devices) to train a one-class Support Vector Machine (SVM) classifier with a Radial Basis Function (RBF) kernel. The trained model was then used on a testing set comprising the collected measurements from the other 50% of the Trojan-free population (i.e., 500 devices) and the entire Trojan-infested population (i.e., 500 devices). Results from a 10-fold cross validation of the above experiment are presented in Table II(a). This confusion matrix reveals that SSCF predicts correctly the Trojan-free and Trojan-infested devices with probabilities of only 0.5013 and 0.3956, respectively. To visualize and better highlight the reason behind this very low classification accuracy, in Figure 14(a) we plot three randomly chosen among the six dimensions of the collected Trojan-free and Trojan-infested signatures (i.e., transmission power). Evidently, the two populations fall upon each other and are inseparable in this 3-dimensional space. We then used Principal Component Analysis (PCA) and, in Figure 14(b), we project the two populations onto three dimensions corresponding to the first three principal components. Once again, the Trojan-free and Trojan-infested populations have significant overlap, explaining why the trained SVM fails to correctly learn an accurate classification boundary. We also repeated the experiment with a 12-dimensional dataset to investigate whether additional features from more fine-grained transmission power levels would provide better discrimination. However, as shown by the confusion matrix in Table II(b) and the PCA plot of Figure 14(c), this is not the case: SSCF still performs poorly.

TABLE III
SVM WITH RBF KERNEL AGAINST ANALOG TROJAN: (A) 6D INPUT DATASET AND (B) 12D INPUT DATASET

(a)

Actual \ Predicted	Clean	Infested
Clean	0.502	0.498
Infested	0.435	0.565

(b)

Actual \ Predicted	Clean	Infested
Clean	0.499	0.501
Infested	0.471	0.529

We followed the same process to also evaluate SSCF against the analog Trojan. Table III(a) shows that for a 6-dimensional dataset the method predicts correctly the Trojan-free and Trojan-infested devices with probabilities of only 0.5022 and 0.565, respectively. Similar to the RF Trojan, visualization of the Trojan-free and Trojan-infested signatures by projection on three randomly chosen dimensions (i.e., raw power measurements) in Figure 15(a) reveals that the populations are overlapping. This overlap remains significant even after using PCA and projecting on the three principal components, as shown in Figure 15(b). The situation remains unchanged when the 12 dimensional dataset is employed, as shown in Figure 15(c) and numerically captured in Table III(b).

In short, since the proposed RF and Analog Trojans do not violate any wireless protocol or circuit specifications, even the most advanced hardware Trojan detection methods, such as SSCF, fail to detect them.

E. Defense Effectiveness

Unlike post-production tests and statistical side-channel fingerprinting, which fail to detect the two hardware Trojan attacks, the proposed ACE defense is able to effectively uncover the Trojan presence. In this section, we evaluate ACE against the two Trojan attacks under various practical operating conditions. Specifically, experiments were carried out under (i) Line of Sight (LoS) communication between the two wireless nodes under various separation distances, (ii) non-Line of Sight (nLoS) communication, where the wireless nodes were placed in adjacent rooms, and (iii) various hardware Trojan operation characteristics, such as data rates and Trojan amplitude levels. Figure 16 and Figure 17 show the corresponding results when evaluating ACE against the RF and the Analog Trojan, respectively, when the legitimate communication operates under BPSK modulation.

1) LoS: The first experiment evaluates ACE under an LoS communication. Figure 16(a) and Figure 17(a) show the output of ACE for a Trojan-free communication. Since ACE removes fading and additive noise from the channel estimates, the resulting heat-map shows minimum variation and remains uniformly colored due to the absence of malicious activity. In the presence of a hardware Trojan, however, the output heat-map shows distinct color variations. Figure 16(b) and Figure 17(b) correspond to the two hardware Trojan-infested communications, where the nodes were placed at a distance of 0.5 m. Here, the Trojan circuits leak information at a rate of 1 bps and the leaked information bits are "0101 1101 1111". Unlike the heat-map corresponding to the Trojan-free case, the ACE output in this case shows distinct color

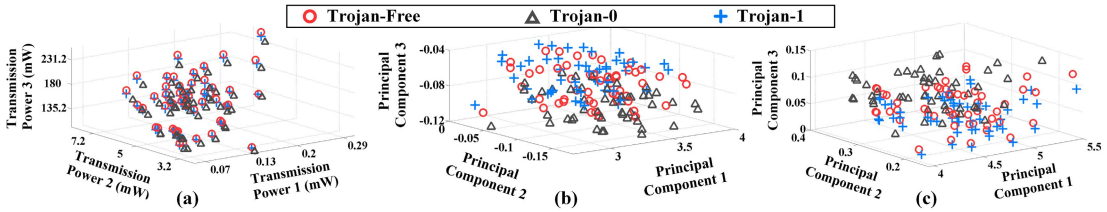


Fig. 14. SSCF for RF Trojan: (a) Trojan-free and Trojan-infested devices projected in a three dimensional transmission power space, (b) output of PCA showing the three maximally variant principal components for a six dimensional data, (c) output of PCA showing the three maximally variant principal components for a twelve dimensional data.

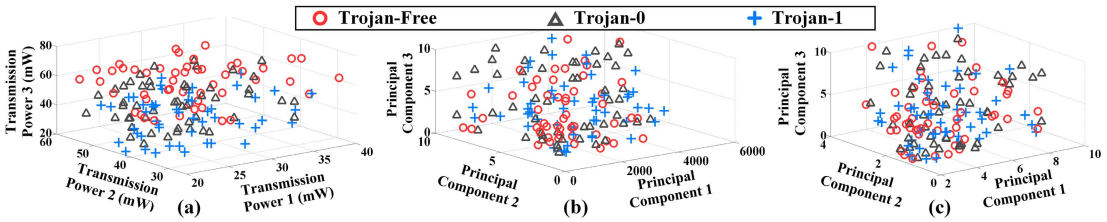


Fig. 15. SSCF for Analog Trojan: (a) Trojan-free and Trojan-infested devices projected in a three dimensional transmission power space, (b) output of PCA showing the three maximally variant principal components for a six dimensional data, (c) output of PCA showing the three maximally variant principal components for a twelve dimensional data.

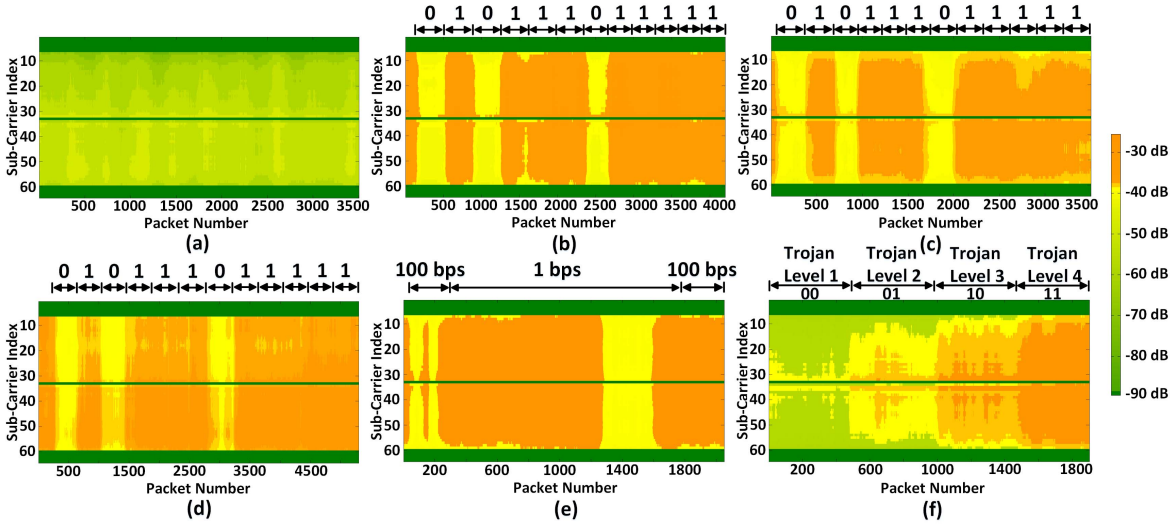


Fig. 16. ACE against RF Trojan for the following communication scenarios: (a) Trojan-free, (b) Trojan-infested over 0.5m, (c) Trojan-infested over 3m, (d) Trojan-infested nLoS, (e) Trojan-infested for varying data rate, and (f) Trojan-infested for different Trojan amplitudes.

variations, revealing Trojan activity. The same experiment was repeated for a node separation distance of 3 m and results are shown in Figure 16(c) and Figure 17(c), respectively. Once again, the distinct color variations in the output heat-map reflect Trojan activity, which is successfully detected by ACE, independent of the node separation distance.

2) nLoS: The next experiment characterizes the effectiveness of ACE in detecting the Trojan activity under a nLoS communication. In this setup, the communicating nodes were placed in adjacent rooms, with a separation distance of 7 m. Similar to the LoS experiment, the Trojan circuits leak information at a rate of 1 bps and the leaked information bits are "0101 1101 1111". Figure 16(d) and Figure 17(d) show the output of ACE for the two hardware Trojan attacks. Despite significant fading and path loss, which is represented by the minor disturbances between the detected Trojan bits in the output heat-maps, ACE is able to successfully reveal the Trojan activity.

3) Varying Trojan Data Rate: The third setup analyzes the effectiveness of ACE against hardware Trojans that dynamically alter their data rate. For example, in this experiment, the Trojan circuits first leak information bits "0101" at a rate of 100 bps, then reduce their rate to 1 bps when leaking "1101" and then increase their rate back to 100 bps when leaking "1111". Figure 16(e) and Figure 17(e) show the corresponding output of ACE for this experiment, where the proposed defense was able to isolate the Trojan activity, even in the presence of unknown channel fading conditions.

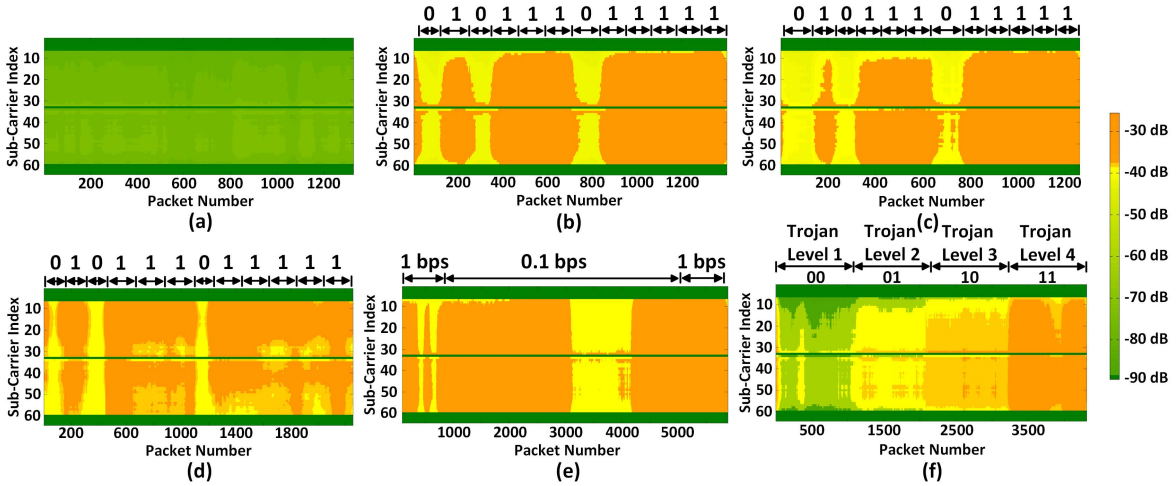


Fig. 17. ACE against analog Trojan for the following communication scenarios: (a) Trojan-free, (b) Trojan-infested over 0.5m, (c) Trojan-infested over 3m, (d) Trojan-infested nLoS, (e) Trojan-infested for varying data rate, and (f) Trojan-infested for different Trojan amplitudes.

4) Varying Trojan Signal Amplitude: The final setup evaluates ACE against Trojans that encode leaked information using multiple Trojan signal amplitudes. For example, instead of leaking one bit at a time, where the Trojan-infested transmitted signal has two different amplitudes, the Trojan can increase its throughput by leaking P bits at a time, which are encoded in 2^P amplitude levels. In this setup, the Trojan circuits leak 2 bits at a time, which are encoded into 4 levels that are 0.5 dB apart. Figure 16(f) and Figure 17(f) show the corresponding results when evaluating ACE against the two Trojan attacks, where the four power levels are revealed using different color representation. The gradient from the smallest to the largest power level verifies the trade-off between rogue reception reliability and hardware Trojan inconspicuousness.

V. Discussion

In this section, we contrast the two analog/RF hardware Trojan attacks and the ACE-based defense with the state-of-the-art in Trojan attacks and defenses in wireless networks.

A. Covert Channel Attacks

Covert channel attacks in wireless networks predominantly exploit vulnerabilities in the baseband logic [7], [9], [27], [28]. The first analog/RF hardware Trojan attack in wireless networks was proposed in [8] and demonstrated in silicon [11] on a wireless cryptographic IC. Here, the Trojan circuit resides within the RF front-end of the device and leaks sensitive information by modulating the amplitude and frequency of the transmitted signal. Similarly, unauthorized leakage of sensitive information below the noise floor of the communication channel was investigated in [10]. Here, the Trojan circuit uses spread spectrum technique to exfiltrate the secret information. However, these attacks were only demonstrated using simple wireless links.

B. Defenses

In [29], an information flow tracking-based proof-carrying hardware solution was proposed for detecting hardware Trojans in wireless devices. While this method is capable of detecting the leakage of sensitive information from the digital domain to the analog/RF front-end during simulation, it falls short in detecting hardware Trojans that are introduced during fabrication. Statistical side-channel fingerprinting is one of the most powerful hardware Trojan detection method found in literature [1], [3], [25], [26]. This method leverages the statistics of side-channel parameters such as power consumption, path delay, supply current, temperature, or combinations thereof, to distinguish between Trojan-free and Trojan-infested devices. In [8], [11], the authors evaluated the hardware Trojan detection capabilities of this method in the context of a wireless cryptographic IC. In [5] and [30], extensions of this technique have been proposed to address the method's reliance on trusted devices (i.e., golden chips) and to detect dormant malicious circuits that get activated in the field, respectively. Similarly, a defense method capable of self-referencing its performance to detect hardware Trojan activity was proposed in [31]. While this technique does not rely on golden ICs, its effectiveness in wireless networks largely depends on the SNR and Trojan-to-circuit ratio. When evaluated against the two hardware Trojans introduced in Section II, these defenses fall short, since the Trojan overhead and the SNR increase required by the Trojan-infested communication are negligible.

C. Comparison

In contrast to prior hardware Trojan attacks, the Trojan examples presented in this work *(i) exploit vulnerabilities in the analog/RF front-end to stage their attack, (ii) are stealthy and yet have higher throughput, thereby constituting a more serious threat, and (iii) have been demonstrated using complex, standards-compliant wireless hardware.* Likewise, in contrast to prior defense techniques, the ACE-based defense proposed and evaluated herein *(i) is implemented on the receiver side, hence its effectiveness cannot be tampered with by an attacker, (ii) does not rely on golden devices (Trojan-free ICs) and (iii) is based on general principles of wireless communication and does not assume any knowledge of the hardware Trojan attack specifics.*

VI. Conclusion

Wireless devices have inherent margins between their operating point and the boundaries defined by their circuit specifications and wireless standards. These margins can be exploited by hardware Trojans to establish covert communication channels and compromise the security of a wireless device. Towards understanding the implications of this threat, we first theoretically analyzed the risk posed by amplitude-modulating analog/RF hardware Trojans in wireless networks. Then, using a WARP-based platform, we experimentally demonstrated the robustness and inconspicuousness of two instances of these malicious circuits in the context of an IEEE 802.11a/g network. Finally, we proposed ACE, a defense method which uses adaptive channel estimation to expose hardware Trojan activity in the presence of channel fading and device noise. The proposed method is implemented on the receiver side, and therefore, its performance cannot be tampered with by an attacker. Experiments conducted in actual channel conditions and over different Trojan amplitudes and data rates verify the effectiveness of the ACE-based defense in detecting hardware Trojans in wireless networks.

References

[1] K. Xiao, D. Forte, Y. Jin, R. Karri, S. Bhunia, and M. Tehranipoor, "Hardware Trojans: Lessons learned after one decade of research," *ACM Trans. Des. Automat. Electron. Syst.*, vol. 22, no. 6, pp. 1–23, 2016.

[2] M. Rostami, F. Koushanfar, and R. Karri, "A primer on hardware security: Models, methods, and metrics," *Proc. IEEE*, vol. 102, no. 8, pp. 1283–1295, Aug. 2014.

[3] M. Tehranipoor and F. Koushanfar, "A survey of hardware trojan taxonomy and detection," *IEEE Design Test Comput.*, vol. 27, no. 1, pp. 10–25, Jan./Feb. 2010.

[4] S. Bhunia, M. S. Hsiao, M. Banga, and S. Narasimhan, "Hardware trojan attacks: Threat analysis and countermeasures," *Proc. IEEE*, vol. 102, no. 8, pp. 1229–1247, Aug. 2014.

[5] Y. Liu, K. Huang, and Y. Makris, "Hardware trojan detection through golden chip-free statistical side-channel fingerprinting," in *Proc. 51st ACM/EDAC/IEEE Design Autom. Conf. (DAC)*, Jun. 2014, pp. 1–6.

[6] L. Lin, W. Burleson, and C. Paar, "MOLES: Malicious off-chip leakage enabled by side-channels," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design-Dig. Tech. Papers*, Nov. 2009, pp. 117–122.

[7] K. S. Subraman, 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 Trans. Inf. Forensics Security*, vol. 14, no. 10, pp. 2720–2734, Oct. 2019.

[8] Y. Liu, Y. Jin, and Y. Makris, "Hardware Trojans in wireless cryptographic ICs: Silicon demonstration & detection method evaluation," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design (ICCAD)*, Nov. 2013, pp. 399–404.

[9] N. Kiyavash, F. Koushanfar, T. P. Coleman, and M. Rodrigues, "A timing channel spyware for the CSMA/CA protocol," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 3, pp. 477–487, Mar. 2013.

[10] D. Chang, B. Bakkaloglu, and S. Ozev, "Enabling unauthorized RF transmission below noise floor with no detectable impact on primary communication performance," in *Proc. IEEE 33rd VLSI Test Symp. (VTS)*, Apr. 2015, pp. 1–4.

[11] Y. Liu, Y. Jin, A. Nosratinia, and Y. Makris, "Silicon demonstration of hardware trojan design and detection in wireless cryptographic ICs," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 25, no. 4, pp. 1506–1519, Apr. 2017.

[12] G. Camurati, S. Poeplau, M. Muench, T. Hayes, and A. Francillon, "Screaming channels: When electromagnetic side channels meet radio transceivers," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2018, pp. 163–177.

[13] G. Goller and G. Sigl, "Side channel attacks on smartphones and embedded devices using standard radio equipment," in *Proc. Int. Workshop Constructive Side-Channel Anal. Secure Design*, 2015, pp. 255–270.

[14] K. S. Subramani, A. Antonopoulos, A. A. Abotabl, A. Nosratinia, and Y. Makris, "ACE: Adaptive channel estimation for detecting analog/RF trojans in WLAN transceivers," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design (ICCAD)*, Nov. 2017, pp. 722–727.

[15] A. Antonopoulos, C. Kapatsori, and Y. Makris, "Hardware Trojans in analog, mixed-signal, and RF ICs," in *The Hardware Trojan War*. Cham, Switzerland: Springer, 2018, pp. 101–123.

[16] X. Cao, Q. Wang, R. L. Geiger, and D. J. Chen, "A hardware trojan embedded in the inverse widlar reference generator," in *Proc. IEEE 58th Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Aug. 2015, pp. 1–4.

[17] C. Cai and D. Chen, "Performance enhancement induced trojan states in op-amps, their detection and removal," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2015, pp. 3020–3023.

[18] Q. Wang, R. L. Geiger, and D. Chen, "Hardware trojans embedded in the dynamic operation of analog and mixed-signal circuits," in *Proc. Nat. Aerosp. Electron. Conf. (NAECON)*, Jun. 2015, pp. 155–158.

[19] R. C. Manso. (2003). *Performance Analysis of M-QAM with Viterbi Soft-Decision Decoding*. [Online]. Available: https://calhoun.nps.edu/handle/10945/1090

[20] *IEEE 802.11-2012 Standard for Information Technology*. Accessed: Nov. 2016. [Online]. Available: https://standards.ieee.org/findstds/standard/802.11-2012.html

[21] J. Proakis and M. Salehi, *Digital Communications*. New York, NY, USA: McGraw-Hill, 2001.

[22] Y. Liu, Z. Tan, H. Hu, L. J. Cimini, and G. Y. Li, "Channel estimation for OFDM," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 1891–1908, 4th Quart., 2014.

[23] F. Peng, J. Zhang, and W. E. Ryan, "Adaptive modulation and coding for IEEE 802.11n," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Mar. 2007, pp. 656–661.

[24] Y. Jin and Y. Makris, "Hardware Trojans in wireless cryptographic ICs," *IEEE Design Test of Comput.*, vol. 27, no. 1, pp. 26–35, Jan./Feb. 2010.

[25] D. Agrawal, S. Baktir, D. Karakoyunlu, P. Rohatgi, and B. Sunar, "Trojan detection using IC fingerprinting," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2007, pp. 296–310.

[26] Y. Jin and Y. Makris, "Hardware trojan detection using path delay fingerprint," in *Proc. IEEE Int. Workshop Hardw.-Oriented Secur. Trust*, Jun. 2008, pp. 51–57.

[27] A. Dutta, D. Saha, D. Grunwald, and D. Sicker, "Secret agent radio: Covert communication through dirty constellations," in *Proc. Int. Workshop Inf. Hiding*, 2012, pp. 160–175.

[28] J. Classen, M. Schulz, and M. Hollick, "Practical covert channels for WiFi systems," in *Proc. IEEE Conf. Commun. Netw. Secur. (CNS)*, Sep. 2015, pp. 209–217.

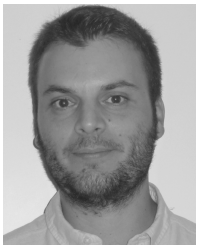
[29] M.-M. Bidmeshki, A. Antonopoulos, and Y. Makris, "Information flow tracking in analog/mixed-signal designs through proof-carrying hardware IP," in *Proc. Design, Autom. Test Eur. Conf. Exhib. (DATE)*, Mar. 2017, pp. 1707–1712.

[30] Y. Liu, G. Volanis, K. Huang, and Y. Makris, "Concurrent hardware trojan detection in wireless cryptographic ICs," in *Proc. IEEE Int. Test Conf. (ITC)*, Oct. 2015, pp. 1–8.

[31] F. Karabacak, U. Y. Ogras, and S. Ozev, "Detection of malicious hardware components in mobile platforms," in *Proc. 17th Int. Symp. Qual. Electron. Design (ISQED)*, Mar. 2016, pp. 179–184.

Kiruba Sankaran Subramani (Member, IEEE) received the B.Tech. degree (Hons.) in electronics and communication engineering from the Amrita School of Engineering, India, in 2009, and the M.S. and Ph.D. degrees in electrical and computer engineering from The University of Texas at Dallas in 2013 and 2018, respectively. He is currently a Post-Doctoral Research Associate in electrical and computer engineering with The University of Texas at Dallas. His research interests include hardware security in wireless networks and design of secure and robust analog/RF integrated circuits and systems. He was a recipient of the Best Hardware Demonstration Award from the 2018 IEEE Symposium on Hardware Oriented Security and Trust (HOST'18).

Noha Helal (Student Member, IEEE) received the B.Sc. degree in electronics and communications engineering from Alexandria University, Egypt, in 2010, and the M.Sc. degree from Nile University, Egypt, in 2012. She is currently pursuing the Ph.D. degree in electrical and computer engineering with The University of Texas at Dallas. Her research interests include information theory and wireless communication and their applications in physical layer security and cooperative communication.



Angelos Antonopoulos (Member, IEEE) received the M.Eng. degree from the School of Electronic and Computer Engineering, Technical University of Crete, Chania, Greece, in 2005, and the M.Sc. and Ph.D. degrees from the School of Electronic and Computer Engineering in 2008 and 2014, respectively. In 2015, he joined the Trusted and RELiable Architectures (TRELA) Research Laboratory, The University of Texas at Dallas (UTD), Richardson, TX, USA, as a Post-Doctoral Research Associate. He is currently a Patent Engineer with u-blox Athens S.A., Greece. His research interests include the design of trusted and reliable analog/RF integrated circuits and systems, hardware security in wireless networks, and design-oriented compact modelling of advanced semiconductor devices. He was a recipient of the Best Hardware Demonstration Award from the 2018 IEEE International Symposium on Hardware Oriented Security and Trust (HOST'18).

Aria Nosratinia (Fellow, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana–Champaign in 1996. He has held visiting appointments at Princeton University, Rice University, and UCLA. He is currently an Erik Jonsson Distinguished Professor and an Associate Head of the Electrical and Computer Engineering Department, The University of Texas at Dallas. His interests lie in information theory and signal processing, with application in wireless communications. He is a fellow of the IEEE for contributions to multimedia and wireless communications. He has received the National Science Foundation Career Award, and the Outstanding Service Award from the IEEE Signal Processing Society, Dallas Chapter. He has served in the organizing and technical program committees of a number of conferences, most recently as a general co-chair of IEEE Information Theory Workshop 2018. He has served as an Editor and an Area Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and as an Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE SIGNAL PROCESSING LETTERS, the *IEEE Wireless Communications* (Magazine), and the *Journal of Circuits, Systems, and Computers*. He was named a highly cited researcher by Clarivate Analytics (formerly Thomson Reuters).

Yiorgos Makris (Senior Member, IEEE) received the Diploma degree in computer engineering from the University of Patras, Greece, in 1995, and the M.S. and Ph.D. degrees in computer engineering from the University of California at San Diego, San Diego, in 1998 and 2001, respectively. After spending a decade on the faculty of Yale University, he joined The University of Texas at Dallas (UT Dallas), where he is currently a Professor of electrical and computer engineering, leading the Trusted and RELiable Architectures (TRELA) Research Laboratory, and the Safety, Security and Healthcare thrust leader for Texas Analog Center of Excellence (TxACE). His research focuses on applications of machine learning and statistical analysis in the development of trusted and reliable integrated circuits and systems, with particular emphasis in the analog/RF domain. He was a recipient of the 2006 Sheffield Distinguished Teaching Award, Best Paper Awards from the 2013 IEEE/ACM Design Automation and Test in Europe (DATE'13) Conference, and the 2015 IEEE VLSI Test Symposium (VTS'15), and the Best Hardware Demonstration Awards from the 2016 and the 2018 IEEE Hardware-Oriented Security and Trust Symposia (HOST'16 and HOST'18). He serves as an Associate Editor for the IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS and has served as an Associate Editor for the IEEE INFORMATION FORENSICS AND SECURITY and the IEEE DESIGN & TEST OF COMPUTERS PERIODICAL, and as a Guest Editor for the IEEE TRANSACTIONS ON COMPUTERS and the IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS.