PyCRA: Physical Challenge-Response Authentication For Active Sensors Under Spoofing Attacks

Yasser Shoukry, Paul Martin, Yair Yona, Suhas Diggavi, and Mani Srivastava
Electrical Engineering Department, University of California at Los Angeles, USA
{yshoukry, pdmartin, yairyo99, suhasdiggavi,mbs}@ucla.edu

ABSTRACT

Embedded sensing systems are pervasively used in life- and security-critical systems such as those found in airplanes, automobiles, and healthcare. Traditional security mechanisms for these sensors focus on data encryption and other post-processing techniques, but the sensors themselves often remain vulnerable to attacks in the physical/analog domain. If an adversary manipulates a physical/analog signal prior to digitization, no amount of digital security mechanisms after the fact can help. Fortunately, nature imposes fundamental constraints on how these analog signals can behave. This work presents PyCRA, a physical challenge-response authentication scheme designed to protect active sensing systems against *physical* attacks occurring in the analog domain. PyCRA provides secure active sensing by continually *challenging* the surrounding environment via random but deliberate physical probes. By analyzing the responses to these probes, the system is able to ensure that the underlying physics involved are not violated, providing an authentication mechanism that not only detects malicious attacks but provides resilience against them. We demonstrate the effectiveness of PyCRA in detecting and mitigating attacks through several case studies using two sensing systems: (1) magnetic sensors like those found on gear and wheel speed sensors in robotics and automotive, and (2) commercial Radio Frequency Identification (RFID) tags used in many security-critical applications. In doing so, we evaluate both the robustness and the limitations of the PyCRA security scheme, concluding by outlining practical considerations as well as further applications for the proposed authentication mechanism.

Categories and Subject Descriptors

C.2.0 [**COMPUTER-COMMUNICATION NETWORKS**]: General: Security and protection

General Terms

Security

CCS'15, October 12–16, 2015, Denver, Colorado, USA.

DOI: http://dx.doi.org/10.1145/2810103.2813679.

Keywords

Embedded Security; Active sensors; Challenge-response authentication; Spoofing attacks; Physical attacks

1. INTRODUCTION

Recent decades have witnessed a proliferation in embedded sensors for observing a variety of physical phenomena. Increased use of these sensors in security- and life-critical applications has been accompanied by a corresponding increase in attacks targeting sensing software, hardware, and even physical, analog signals themselves. While considerable research has explored sensor security from a system-level perspective—network redundancy, sensor fusion, and the like—sensors themselves remain largely vulnerable to attacks targeting analog signals prior to digitization. This vulnerability can lead to catastrophic failures when a malicious third party attempts to spoof the sensor [19, 14, 3, 33].

Several *system-level* sensor security schemes have been proposed in the context of power grids. For example, Dorfler et al. have explored distributed cyber-physical attack detection in the context of power networks [6, 30]. Similar ideas for providing system-level security in smart grids can be found in [16, 18, 4, 22, 35]. Security schemes in this vein include, among others, state-space and control-theoretic approaches for detecting anomalous system behavior [7, 30]. One idea common to these efforts is that an inherent security mechanism and robustness can be found in the physics governing the dynamics of the *system* as a whole. For example, a mismatch between the rate of change in a vehicle's location as reported by GPS and by the odometer sensor may indicate that one of these two sensors is either faulty or under attack.

A complementary security mechanism can be found in the underlying physics governing the *sensor* itself. If a sensor observes an analog signal that appears to violate the physics governing the sensing dynamics, the signal itself may be under attack, necessitating security mechanisms at the analog signal level. To reduce sensor-level vulnerabilities, engineers often place sensors in secure or remote physical locations to preclude direct physical contact with the sensing hardware. Additionally, the phenomenon being sensed is often difficult to access, whether prohibitively far away or surrounded by protective material. In such scenarios, adversaries have access only to the analog signal prior to it reaching the sensor, and their attack must be carried out without direct access to any hardware in the entire sensing path, from source to sink. Even with these countermeasures in place, an adversary can still attack sensors by manipulating the physical signals before their transduction and subsequent digitization [19, 33]. Robust countermeasures for such attacks must necessarily be carried out at the physical level as well—once these signals have been sampled and digitized, no amount of post-processing can repair the compromised sensor data.

Broadly speaking, sensors can be divided into two categories: passive (those that sense pre-existing physical signals) and active (those that perform some action to evoke and measure a physical response from some measurable entity). Examples of passive sensors include temperature, humidity, and ambient light, while active sensors include ultrasound, laser scanners, and radar. Passive sensors are largely naïve listening devices–they blindly relay information to higher levels of software without regard for the integrity of that information. Digital filtering and other post-processing techniques can be used to remove noise from passive sensors, but they remain unable to combat attacks at the physical layer in any meaningful way. On the other hand, active sensors introduce the possibility for more advanced security measures. PyCRA is, at its core, a method of ensuring the trustworthiness of information obtained by active sensors by comparing their responses to a series of physical queries or challenges. The driving concept behind PyCRA is that, by periodically stimulating the environment with a known signal and measuring the response, we can ensure that the signal measured by the sensor is in accordance with the underlying sensing physics. This periodic stimulation and subsequent behavioral analysis—the physical challenge-response authentication, creating a secure active sensing platform—is the main contribution of this work.

We demonstrate the effectiveness of PyCRA for three exemplary cases: physical attack detection for magnetic encoders, physical attack resilience for magnetic encoders, and passive eavesdropping detection for RFID readers. Magnetic encoders are used in a wide array of commercial and industrial applications and are representative of a large class of inductive active sensors. We demonstrate not only how active spoofing attacks can be detected for these inductive sensors but also how the effects of these attacks can be mitigated. Eavesdropping detection on RFID readers serves to illustrate an extension of PyCRA to enable detection of *passive* attacks. Our results from more than 90 experiments demonstrate that PyCRA can accurately detect attacks in a variety of settings. We believe that the methods demonstrated in this work can be applied to a broad array of active sensors beyond those studied directly in this work, including ultrasound, optical sensors, active radar, and more.

1.1 Contributions of PyCRA

In summary, the contributions described in this paper are threefold:

- We present a generalizable physical challenge-response authentication scheme for active sensing subsystems.
- We introduce algorithms for detecting the presence of and providing resilience against physical attacks when using physical challenge-response authentication.
- We demonstrate the effectiveness of PyCRA, our implementation of physical challenge-response authentication, against several different attack types and for over 90 experiments with three exemplary applications: (1) detection of active attacks on magnetic encoders, (2) resilience against active attacks on magnetic encoders, and (3) detecting passive eavesdropping attacks on RFID readers.

The rest of this paper is organized as follows. Section 2 outlines the attacker model. Section 3 describes the basic operation of the PyCRA authentication scheme for detecting active attacks. Section 4 outlines theoretical limitations of attackers on physical signals. Section 5, 6, and 7 are devoted to the results of three case studies: attack detection for magnetic encoders, attack detection for passive eavesdropping on RFID readers, and attack resilience for magnetic encoders. Finally, we offer a discussion and concluding thoughts in Sections 8.1, 8.2 and 9.

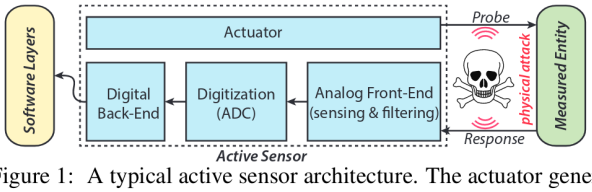


Figure 1: A typical active sensor architecture. The actuator generates an analog signal (energy) which is reflected by the measured entity back to the sensor. The received analog signal is captured and processed by the analog front-end. The signal is then converted to a digital format which is processed once more (by the digital back-end) before being sent to higher level software layers.

2. ATTACKER MODEL

Before describing mechanisms by which we can detect and prevent sensor attacks at the physical layer, we must differentiate between two broad categories of sensors—namely passive and active sensors—and define what we mean by a physical attack.

2.1 Passive vs. Active Sensors

Sensors can be broadly classified as either passive or active based on the source of energy being sensed. Passive sensors measure ambient energy. For example, temperature sensors like those found in thermostats are considered passive, because they measure heat energy in the ambient environment. By contrast, active sensors probe some physical entity with self-generated energy as shown in Figure 1. This energy is partially reflected back to the sensor where it is measured and used to infer properties about some physical phenomenon. Examples of active sensors include ultrasonic range finders (used in robotics), optical and magnetic encoders (used in automotive vehicles, industrial plants, & chemical refineries), radar, and even radio-frequency identification (RFID) systems. In RFID, a reader is used to generate electromagnetic waves which are then used by wireless tags to transfer back their unique identifier to the reader.

In this paper, we focus on providing security for active sensors. In particular, we leverage an active sensor's ability to emit energy in order to 1) provide detection of active attackers trying to spoof the sensor, 2) mitigate the effects of active spoofing attacks and 3) detect passive eavesdropping attacks attempting to listen to the information received by the sensor. In the following subsections, we define what we mean by physical attacks on active sensors and outline the assumed properties and limitations of a potential adversary.

2.2 Defining Physical Attacks

In this paper, a physical attack refers to a malicious alteration of a physical, analog signal (e.g., magnetic waves, acoustic waves, visible waves) prior to transduction and digitization by a sensor, as shown in Figure 1.

2.3 Adversarial Goals

The adversary considered in this work has a number of goals related to misinforming and misleading sensors. These goals are summarized below.

G1 *Concealment: An attacker does not want the presence of his or her attack to be known.*

If a sensor attack can be easily detected, preventative countermeasures like hardware redundancy and resilience at the system-level can often be used to mitigate the damage done by the attack [7, 30].

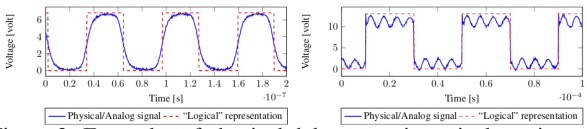


Figure 2: Examples of physical delays seen in typical sensing and actuation hardware, including optical sensors (left) and electromagnetic coupled (e.g., RFID) sensors (right). In each case, the measured analog signal (blue solid) lags behind the ideal, "logical" signal (red dashed), causing delays.

G2 *Signal Injection: An attacker will attempt to trick the sensor into thinking that a malicious, injected signal is the true physical signal.*

The primary goal of an attack is to replace the true physical signal that a sensor aims to sense with a malicious signal. In other words, an adversary will attempt to "inject" a signal into the physical medium that the sensor is measuring in order to jam or spoof the sensor.

G3 *Signal Masking: An attacker will attempt to prevent the sensor from being able to detect the true physical signal.*

If the sensor is still capable of reliably discerning the correct signal from the malicious, injected signal, then the attack may not be successful. Thus, the adversary aims not only to inject a signal but also to mask the true signal, whether by overpowering, modifying, or negating (canceling) it.

2.4 Assumptions about the Adversary

The physical attacks against sensors considered in this work operate under four main assumptions:

A1 *Non-invasiveness: Attacks are of a* non-invasive *nature—that is, the attacker is not allowed direct access to the sensor hardware. Additionally, the adversary does not have access to the sensor firmware or software, whether directly or through wired or wireless networking.*

In most life- and safety-critical applications, engineers are careful to ensure that sensors are not physically exposed and vulnerable to direct tampering. For example:

- Sensors are often installed inside the body of a physically secured infrastructure (e.g., sensors inside the body of an automotive system, moving UAV drones, etc.).
- For sensors which are physically accessible, existing techniques in the literature demonstrate ways to implement tamper-proof packaging to protect sensors from direct, physical modifications [31, 17, 1].
- Numerous sensor systems have methods for detecting when wires connecting their various sensors have been tampered with. For example, automotive systems are equipped with sensor failure detection systems which can detect whether all sensor subsystems are correctly connected and alert the driver if any of them fails [8].

Because of this, any attack must be carried out from a distance, without direct access to any sensor hardware. In short, an adversary is assumed to have access only to the physical/analog medium used by the sensor—magnetic waves, optics, acoustics, etc.

Additionally, it is important to distinguish these sensors from *sensor nodes* (which appear in the literature of sensor networks); the attacks and countermeasures in this work target *sensors* themselves. Sensors are simple subsystems designed to perform only one simple task; sensing the physical world. Because of this, many sensors do not support remote firmware updates and do not typically receive commands from a remote operator, making such attack vectors uncommon as many sensors do not have such capabilities.

A2 *Trusted Measured Entity We assume that the physical entity to be measured by the sensor is trusted and incapable of being compromised.*

Similar to the sensor hardware itself, the entity that the sensor aims to measure is typically difficult to access or alter directly while maintaining Goals G1–G3. For example, in RFID systems the tag itself is often encased in tamper-proof packaging [31, 17]; for ultrasonic ranging and active radar, maliciously altering the measured entity (often the entire surrounding environment) is impractical in time & effort and undoubtedly violates Goal G1; for airplane engine speed sensors, the engines cannot easily be modified or replaced; for heart monitors, the heart cannot (we hope) be modified [19], and so forth.

A3 *Physical Delays (τ_{attack}): Adversaries require physical hardware with inherent physical delays. This delay, though variable in duration, is fundamental to all physical actuation and sensing hardware.*

These same analog/physical signals cannot be manipulated or even observed (i.e. sniffed) without physical hardware. That is, to tamper with magnetic waves, an attacker needs hardware that is able to generate magnetic waves, optical signals need physical hardware that generates optical signals, and so on. Furthermore, this hardware has to obey fundamental physics imposed by nature; the underlying physics dictate that the response of any physical element is governed by a dynamical model (mathematically modeled using differential/difference equations) [9, ch. 2], [2, chs. 8–9]. This dynamical model describes the output response for each physical element in response to their inputs, e.g., the time for a voltage to drop from a certain value to zero and so on. Although from a system point of view, we often assume that analog signals like those in Figure 2 take on logical values of 0 and 1, the underlying physics is always different from this "system" point of view. For example, Figure 2 shows how hardware that generates clock waveforms and optical pulse signals behaves quite differently from the desired, logical signals used to control them. In general, no physical signal can arbitrarily jump from one state to another without suffering from *delays* imposed by physics [9, ch. 2].

Furthermore, these physical delays are lower bounded by a non-zero, fundamental limit. For example, the time response of an electromagnetic sensor/actuator is a multiple of physical constants like magnetic permeability [2, chs. 8–9] or permitivity and electric constants for capacitive sensors [9, ch. 4]. In general, the time response of any sensor or actuator can never be below certain fundamental thresholds controlled by physical constants. We refer to this physical delay as τ_{attack} for the remainder of this paper.

A4 *Computational Delays: PyCRA is designed and analyzed with a focus on* physical *delays. We make no assumption regarding the computational power of a potential adversary.*

We assume that an adversary has knowledge of the underlying security mechanism, attempting to conceal an attack by reacting to each physical challenge or probe from the PyCRA-secured active sensor. In practice, such an adversary would suffer from *computational delays* in addition to the physical delays addressed above. These delays would make it even more difficult for an adversary to respond to these challenges in a timely manner. PyCRA is designed

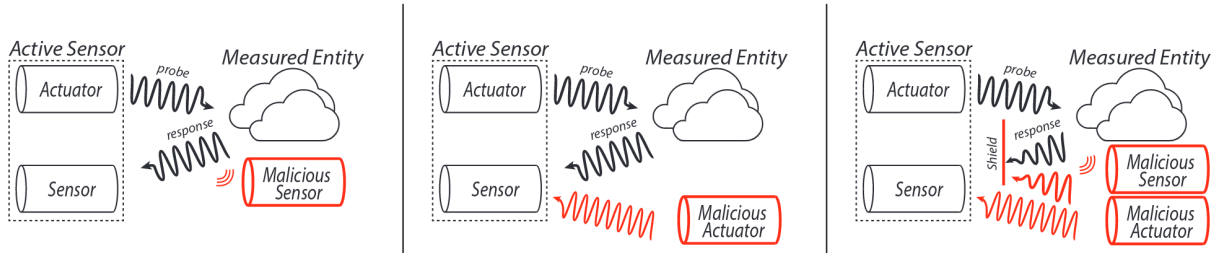


Figure 3: An illustration of three physical attack types: (a) a passive eavesdropping attack, (b) a simple spoofing attack where a malicious actuator blindly injects a disruptive signal, and (c) an advanced spoofing attack where an adversary uses a sensor to measure the original signal and an actuator to actively cancel the original signal and inject a malicious one.

to leverage only the physical delays addressed above, but additional computational delays would make it even easier to detect the presence of an attack.

2.5 Physical Attack Types for Sensors

Attacks can be classified as either passive (eavesdropping) or active (spoofing). While we consider only physical/analog attacks in accordance with assumptions A1–A4, the passivity of an attack is decided by whether or not the attacker is manipulating (or spoofing) the physical signal or merely listening to it. Active attacks themselves can be classified once more into simple spoofing or advanced spoofing attacks. In short, physical sensor attacks in accordance with assumptions A1–A4 can be broadly divided into three categories (Types):

T1 *Eavesdropping Attacks: In an eavesdropping attack, an adversary uses a malicious sensor in order to listen to the active sensor's "communication" with the measured entity (Figure 3a).*

T2 *Simple Spoofing Attacks: In a simple spoofing attack, an adversary uses a malicious actuator to blindly inject a malicious signal in order to alter the signal observed by the sensor. These attacks are simple in that the malicious signal is not a function of the original, true signal (Figure 3b).*

T3 *Advanced Spoofing Attacks In an advanced spoofing attack, an adversary uses a sensor in order to gain full knowledge of the original signal and then uses a malicious actuator to inject a malicious signal accordingly. This enables an attacker to suppress the original signal or otherwise alter it in addition to injecting a malicious signal (Figure 3c).*

We argue that these attack types span all possible modes of attacks that abide by Assumptions A1–A4 with those goals outlined in G1–G3. For example, jamming or Denial of service (DoS) attacks falls in category T2 where the attacker's actuator is used to blindly generate high amplitude, wide bandwidth signals to interfere with the physical signal before it reaches the sensors; replay attacks fall in either category T2 or T3 based on whether the attacker is blindly replaying a physical signal or destructing the original physical signal before inserting the replay signal; spoofing attacks like those demonstrated in [19] fall in category T2; and attacks described in [33] fall within both T2 and T3.

At first glance, attacks of type T1 may not seem important especially if the sensor under attack measures a physical signal that is publicly accessible (e.g., room temperature, car speed, etc.). In such cases, an adversary can measure the same physical signal without the need to "listen" to the interaction between the active sensor and the environment. However, this may not always be the case. For example, an attacker might measure magnetic waves during an exchange between an RFID reader and an RFID tag, learning potentially sensitive information about the tag. These attacks are passive, meaning that the attacker does not inject any energy into the system. Sections 5 describes methods for detecting attack types T2 and T3, leaving attack type T1 for later discussion in Section 6.

3. PYCRA AUTHENTICATION SCHEME

The core concept behind PyCRA is that of physical challenge-response authentication. In traditional challenge-response authentication schemes, one party requires another party to prove their trustworthiness by correctly answering a question or *challenge*. This challenge-response pair could be a simple password query, a random challenge to a known hash function, or other similar mechanisms. In the proposed physical challenge-response authentication, the challenge comes in the form of a *physical* stimulus placed on the environment by an active sensor. Unlike traditional schemes, the proposed *physical* challenge operates in the analog domain and is designed so that an adversary cannot issue the correct response because of immutable physical constraints rather than computational or combinatorial challenges.

We begin by modeling the problem of detecting physical sensor attacks as an authentication problem. To draw this analogy, let us consider the communication system shown in Figure 4a. This figure shows two 'parties': (1) an active sensor composed of actuation and sensing subsystems and (2) the measured entity which responds to signals emitted by the actuator contained within the active sensor. The first party—the active sensor—is responsible for initiating the "communication" by generating some physical signal such as a magnetic, acoustic, or optical wave. The second party—the measured entity—responds to this "communication" by modulating this signal and reflecting it back to the sensing subsystem of the active sensor. With this analogy in mind, the problem of detecting physical attacks can be posed as that of ensuring that the "message" seen by the sensor has originated from a trusted party (the true entity to be measured). This is akin to identity authentication in the the literature of computer security but applied to the analog domain.

3.1 Simple PyCRA Attack Detector

Using the communication analogy shown in Figure 4a and recalling that we are interested only in active sensors as described in Section 2.1, we notice that the measured entity, as a participating party in this communication, is strictly *passive*, i.e. it cannot initiate communication; it responds only when the sensor generates an appropriate physical signal.

PyCRA exploits this "passivity" in order to facilitate the detection of attacks. Without PyCRA, an active sensor's actuator would probe the measured entity in a normal fashion using a determin-

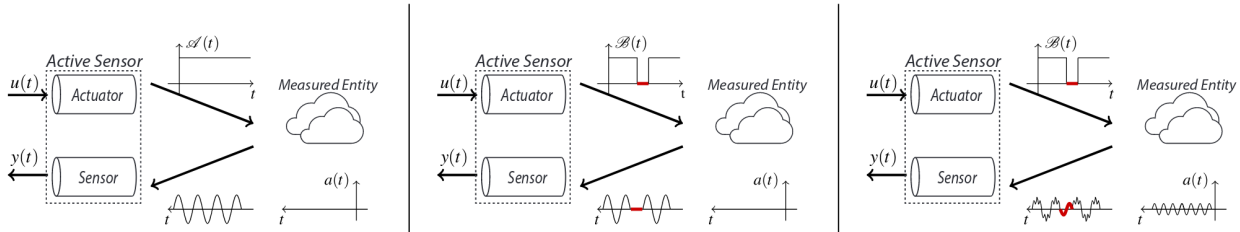


Figure 4: An illustration of the PyCRA architecture and attack detection scheme: (a) During normal operation, the active sensor generates a signal $\mathscr{A}(t)$. This signal passes through environmental dynamics and is reflected back to the sensor as $y(t)$; (b) Using the proposed PyCRA scheme, the sensor generates a modulated signal $\mathscr{B}(t)$. If there is no attack present, the reflected signal diminishes if the active sensor's actuator is driven to zero; (c) Using the proposed PyCRA scheme while the sensor is under attack (by signal $a(t)$), a malicious signal is detected during the period when the actuator is disabled.

istic signal denoted by $\mathscr{A}(t)$. We embed in this signal a physical challenge through pseudo-random binary modulation of the form:

$$\mathscr{B}(t) = u(t)\mathscr{A}(t), \quad u(t) \in \{0,1\} \tag{1}$$

where $u(t)$ is the binary modulation term and $\mathscr{B}(t)$ is the modulated output of the actuator. The output of the active sensor is denoted by $y(t)$ as shown in Figure 4. In the absence of an attacker and from the passivity of the measured entity, setting $u(t) = 0$ (and consequently $\mathscr{B}(t) = 0$) at time $t_{challenge}$ will cause $y(t)$ to go to zero.

Potential attackers must actively emit a signal $a(t)$ to overpower or mask $y(t)$ (Goals G2–G3). A naïve attacker might continue to emit this signal even when $\mathscr{B}(t) = 0$ as shown in Figure 4c. In this case, the attack can be easily detected, since any nonzero $y(t)$ while $u(t) = 0$ can be attributed to the existence of an attacker.

More advanced attackers might attempt to conceal their attacks when they sense the absence of $\mathscr{B}(t)$ as in Goal G1. Due to Assumption A3, an attacker could drive $a(t)$ to zero only after a delay of τ_{attack}, where $\tau_{attack} \geq \tau_{physical\ limit} > 0$ is the unavoidable physical delay inherent in the attacker's hardware. Therefore, the mechanism described above can still detect the presence of an attack within this unavoidable time delay. Furthermore, an attacker cannot learn and compensate for this inherent delay preemptively due to the randomness of the modulation term $u(t)$. Again, any nonzero $y(t)$ sensed while $u(t) = 0$ can be attributed to the existence of an attacker. The simple PyCRA attack detector can be summarized as follows:

[Step 1] Select a random time, $t_{challenge}$
[Step 2] Issue a physical challenge by setting $u(t_{challenge}) = 0$
[Step 3] If $y(t_{challenge}) > 0$, declare an attack

Note that the previous process needs to happen within small amount of time (e.g., in the order of milliseconds) such that it does not affect the normal operation of the system.

3.2 The Confusion Phase

Every physical signal is subject to random perturbations known as noise. A fundamental characteristic of this noise is the *signal to noise ratio* (SNR). This SNR determines the ability of any sensor to distinguish between changes in a signal of interest and the random noise. As with the physical delay, this SNR is fundamental, and it is never equal to zero. As a result, if a signal is within the noise floor (less than the noise amplitude), it is fundamentally impossible to detect any change in the physical signal [36].

As with the physical time delay τ_{attack}, we use this fundamental limit in order to enhance PyCRA and introduce additional security. To do so, we modify the physical challenge by introducing an intermediate step—between the active phase (e.g., $u(t) = 1$) and the silent phase (e.g., $u(t) = 0$)—called the *confusion* phase. In this phase, the active sensor uses its actuator to generate a signal $u(t)$ that is small enough to barely exceed the noise level. Next, we wait in this confusion phase for a random time $t_{confusion}$ before entering the silent time. This process is summarized in Figure 5.

Recall that one of the attacker's goals is to remain stealthy (Goal G1). If the attacker is unable to instantaneously detect the changes in the physical challenge, he or she will reveal themselves. Due to the existence of noise, no attacker—whether using software or hardware to counter the physical challenges issued by PyCRA—can instantaneously detect the change in the physical challenge. That is, there always exists a non-zero probability of the attacker missing the changes in the physical challenge. In Section 4, we detail a theoretical result that explains the relationship between the amplitude of the physical challenge within the confusion phase and the probability that the attacker will fail to detect changes in the physical challenge.

3.3 Effect of Physical Delays at the Sensor

As with the attacker, the actuator used by the active sensor itself suffers from physical delays. This means that when PyCRA issues a physical challenge, the actuator output does not transition immediately. Apparently, if the physical delay in the active sensor is greater than τ_{attack}, then an adversary can conceal his signal. To counter this, PyCRA constructs a mathematical model for the sensor that is used—in real time—to predict and eliminate the effects of the active sensor's physics. By calculating the residual between the expected output and the measured output, PyCRA can still detect the existence of an attack even if the sensor's dynamics are slower than those of the attacker.

3.4 χ^2 PyCRA Attack Detector

If we obtain an accurate model of the sensor's actuator dynamics, then we can remove its effects from the measured response, ensuring that any residual energy measured while $u(t) = 0$ belongs to an external source such as an attacker.

3.4.1 Obtaining the Sensor Model

To compensate for the actuator dynamics, we first need to acquire an accurate model that captures the underlying physics of the active sensor. Below we model the active sensor using the generic nonlinear state update of the form:

$$x(t+1) = f(x(t), u(t)) + w(t) \tag{2}$$

$$y(t) = h(x(t)) + v(t) \tag{3}$$

where $x(t) \in \mathbb{R}^n$ is the active sensor state at time $t \in \mathbb{N}_0$ (e.g., the electrical current and voltages inside the sensor at time t), $u(t) \in \mathbb{R}$ is the modulation input to the sensor, the function $f : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$

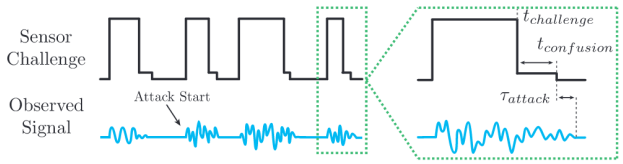


Figure 5: Sensor actuator output (top) with confusion and silence phases and the corresponding raw signal (bottom) with an attack.

is a model describing how the physical quantities of the sensor evolve over time, and the function $h : \mathbb{R}^n \to \mathbb{R}$ models the sensor measurement physics. Such models can be either derived from first principles [9, 2, 10] or through experimental studies [23, 20]. Additionally, these models are used to design the sensors themselves and are typically known to the sensor manufacturers. Finally, since no mathematical model can capture the true system behavior exactly, the term $w(t) \in \mathbb{R}^n$ represents the mismatch between the true sensor and the mathematical model while $v(t)$ models the noise in the sensor measurements.

3.4.2 χ^2 Detector

We use the dynamical model of the sensor (Equations (2) and (3)) in designing a χ^2 detector to detect the existence of an attacker. χ^2 detectors appear in the literature of automatic control, where they are used in designing fault tolerant systems [26, 28, 25, 38]. The χ^2 detector works as follows:

[Step 1] Select random times, $t_{challenge}$ and $t_{confusion}$.

[Step 2] Issue a physical challenge by entering the confusion phase at time $t_{challenge}$ and then enter the silent phase at time $t_{challenge} + t_{confusion}$.

[Step 3] Residual Calculation: Here we use Equations (2) and (3) to calculate an estimate for the current sensor state $\hat{x}(t)$ and the predicted output $\hat{y}(t) = h(\hat{x}(t))$. This operation is initiated at $t_{challenge} + t_{confusion}$ when $u(t)$ transitions to 0—the actuator "silence time"—and terminates once $u(t)$ transitions back to one, signaling the end of actuator "silence."

The model represented by Equations (2) and (3) describes the output of the sensor when the attack is equal to zero. Therefore, the residual[1] between the measured output and the predicted output, $z(t) = y(t) - \hat{y}(t)$, corresponds to both the attack signal as well as the environmental dynamics during the time interval before $u(t)$ drops to 0. For each segment of length T where $u(t) = 0$, we calculate the norm of the residual $z(t)$ as:

$$g(t) = \frac{1}{T} \sum_{\tau=t-T+1}^{t} z^2(\tau) \tag{4}$$

[Step 4] Detection Alarm: Once calculated, we compare the χ^2 residual $g(t)$ against a pre-computed alarm threshold α. This alarm threshold is chosen based on the noise $v(t)$. Whenever the condition $g(t) > \alpha$ is satisfied, the sensor declares that an attacker has been detected.

4. THEORETICAL GUARANTEES

As discussed before, PyCRA is based on the concept that physics impose fundamental and immutable constraints on how quickly an attacker can detect changes in the physical-challenge and how fast he can react to these challenges. In this section, we show a theoretical result that allows PyCRA to increase the probability of an attacker failing to detect the changes in the physical-challenge by correctly designing the confusing phase (discussed in Section 3.2) and hence increase the probability of detecting the attack.

THEOREM 1. *Consider an attacker attempting to detect changes in a physical challenge signal with mis-detection probability α. For any strategy the attacker chooses, and because of the SNR exists at any sensor, the probability of the attacker having a constant detection delay $\tau > 0$ is bounded away from zero, i.e., with high probability the attacker will detect a change and turn off his signal only after time T after the beginning of the confusion period. In addition, decreasing the amplitude of the signal emitted by the active sensor during the confusion period by a factor of $\beta > 1$ increases the delay τ by a factor of β^2.*

PROOF SKETCH. We base the proof on the results reported in [36] on change point detection which measure fundamental limits on checking changes in noisy signal. In the change point detection setting, the false alarm probability is analogous to the event where the attacker switches off his signal before the beginning of the silent period. Delay in [36] is defined as the time that elapses from the change point until the change is detected. When $\alpha \ll 1$, the false alarm probability induces a probability which is proportional to α for the event that change is detected within a time interval shorter than T (a constant independent of α).

Decreasing the amplitude of the signal actuated by the active sensor during the confusion period by a factor of β leads to a decrease in SNR by a factor of β^2. Based on this relation, the attacker has to increase the delay by a factor of β^2 in order to maintain false alarm probability α. □

5. CASE STUDY (1): DETECTING ACTIVE SPOOFING ATTACKS FOR MAGNETIC ENCODERS

Magnetic encoders are active sensors used in a wide array of industrial, robotics, aerospace, and automotive applications. The goal of an encoder is to measure the angular velocity or position of a gear or wheel in order to provide feedback to a motor controller. The operation of these systems depends heavily on the accuracy and timeliness of the individual encoders. This section describes the basic operation of magnetic encoders in particular and the types of attacks that can be mounted against them as well as how PyCRA can be used to provide security for them.

5.1 Magnetic Encoders

Magnetic encoders rely on magnetic variations to measure the angular velocity of a gear or wheel and are often designed to handle dust, mud, rain, and extreme temperatures without failing. The goal of each encoder is to provide a signal whose frequency corresponds to the speed of a gear. These signals are conditioned and passed to a motor controller unit which detects if any corrective actions need to be taken.

Typical magnetic encoders operate by generating a magnetic field in the presence of a rotating ferromagnetic *tone ring* or *tone wheel*. This ring has a number of teeth on its edge so that the reflected magnetic wave as observed by the encoder varies over time as a (noisy) sinusoidal wave. By measuring the frequency of this reflected signal over time, each sensor and consequently the motor controller is able to infer the angular velocity of any given gear, wheel, or motor as illustrated in Figure 6.

Attacks on magnetic encoders have been studied in [33] in the context of Anti-lock Braking Systems in automotive vehicles. Both

[1]The name of the Chi-squared (χ^2) detector follows from the fact that, in the case of no attack, the residual $z(t)$ is a Gaussian random variable, and hence its square $g(t)$ is a χ^2 distributed random variable.

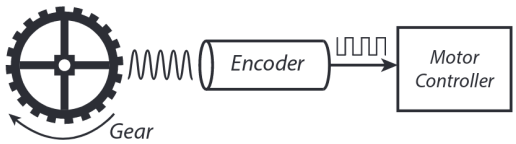


Figure 6: Flow diagram for a typical magnetic encoder: The signal begins as a reflected magnetic wave from a gear. This signal is captured by a pick-up coil or Hall Effect sensor, conditioned into a clean square wave, and finally translated into an angular velocity.

simple spoofing [T2] and advanced spoofing [T3] attacks are shown to influence the vehicle stability. In this case study, we show how PyCRA can detect the existence of such attacks.

5.2 The PyCRA-secured Magnetic Encoder

Physically, the proposed secure magnetic encoder sensor consists of two main parts: (i) the front-end containing the actuator and pickup coils responsible for both probing the rotating tone ring and measuring the response, and (ii) the processing backend. Figure 7 shows the front-end of the sensor used in our evaluation. The actuator coil depicted is much larger than would be required in a commercial product, because it consists of a magnetic core and a hand-wound high-gauge wire. The following is an overview of the main blocks of the sensor.

5.2.1 Actuator Coil

The main component required for the secure sensor is the actuator coil. In this work, we use an insulated copper wire wrapped around a ferromagnetic core and driven using a power amplifier.

5.2.2 Pickup and Filtering

The pickup (measurement) coil is wrapped around the same ferromagnetic core used for the actuator coil. In order to reduce the effect of noise from other EMI sources within the vehicle body, the output of the pickup coil is connected to a differential amplifier with high common-mode rejection. The output of this differential amplifier is connected to the digital processing backend.

Another security concern of the magnetic encoder is the wires connecting the coils to the digital backend. These wires pose a potential vulnerability, as an attacker can cut them and connect his attack module directly. However, such attacks are already accounted for in many systems as addressed in Assumption A1.

5.2.3 Processing Elements

The secure sensor requires enough processing power to perform the necessary computations in real-time. The DSP calculations take place on a high power ARM Cortex (M4 STM32F407) processor, which has ample floating point support. We do not consider any power consumption issues in our design.

5.3 Obtaining the Sensor Model

The dynamics of the sensor (including the actuator, high gain current amplifier, sensors, and the signal conditioning circuit) are identified using standard system identification methods [23]. That is, we applied four different pseudo random binary sequences (PRBS) to the system, collected the output, and then applied subspace system identification techniques in order to build models of increasing complexity [23]. Finally we used both whiteness tests and correlation tests to assess the quality of the obtained model [20]. In order to validate the model, we generated a random sequence similar to those used in the real implementation of the sensor. We fed the same input to both the sensor and the model and recorded the error.

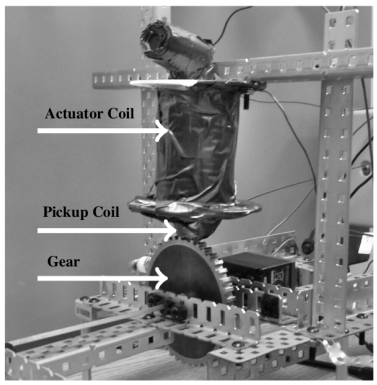


Figure 7: PyCRA encoder actuator coil, sensor, and gear setup.

Experiments show that the model is reasonably accurate with an error in the range of 5 milli-Volts.

5.4 Testbed

In order to test the PyCRA-secured magnetic encoder, we constructed a testbed consisting of the proposed secure sensor attached to a Mazda Rx7 tone ring. The tone ring is attached to a DC motor which simulates a rotating wheel. An additional coil is added to simulate the effect of an attacker. The attacker coil is also controlled by a high gain amplifier controlled through a real-time xPC Target system connected to MATLAB.

A Mazda RX7 magnetic encoder sensor is also attached to the same tone ring in order to provide ground truth. The output of this sensor is connected to a MAX9926U evaluation kit which includes an interface capable of converting the raw sinusoidal wave into the encoded square wave as shown in Figure 6. The output of the proposed secure sensor as well as the output of the MAX9926U is monitored by the same real-time xPC Target for comparison.

5.5 Calibration against natural variations

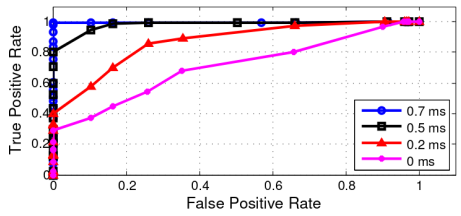
Sensor modeling is usually done in a controlled environment. However, once the sensor is placed in a testbed, multiple natural variations, mechanical asymmetries, and other environmental factors degrade the accuracy of such models. To account for these variations, we use a simple learning mechanism to estimate the noise level in the measurements and the deviation between the expected outputs (as per the model) and the actual outputs. Once these parameters are learned, we can set the alarm threshold accordingly. Results can be further improved by considering online identification-and-calibration of the sensor model.

5.6 Attack Detection for Magnetic Encoders

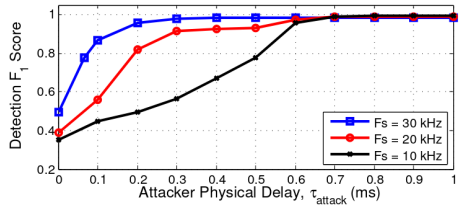
We begin with a simple spoofing attack [T2] in which an attacker injects a sinusoidal wave of varying frequency. Spoofing attacks of this nature attempt to overpower the true frequency of the system and force the sensor to track the false frequency (mirroring the simplistic spoofing attack in [13]). In this experiment the original tone ring frequency is fixed at 71 Hz, and the frequency of the attacking coil increases linearly from 60 Hz to just over 400 Hz.

As per our attacker model in Section 2, we assume that the attacker attempts to conceal his or her presence (Adversarial goal [G1]). This means that the adversary will be able to detect when the actuator coil is turned off and will, after some time τ_{attack}, temporarily halt the attack.

The stealthiness of the attacker necessitates that the PyCRA detection scheme have high accuracy even when the attacker is quick



(a)



(b)

Figure 8: Results from 30 experiments showing (a) the accuracy of attack detection for a simple spoofing attack with sampling rate $F_s = 10$ kHz and a range of τ_{attack}, and (b) attack detection accuracy as a function of τ_{attack} for several sampling rates, F_s.

to react. We evaluated the PyCRA detection scheme across a range of τ_{attack} values, χ^2 detection thresholds (α), and sampling frequencies (F_s). Note that in order to simulate an attacker with 0 ms physical delays (which is physically impossible), we gave the attacker access to the random signal generated by PyCRA so that the attacker can start shutting down his actuators *before* PyCRA generates the physical challenge.

In total, we conducted over 30 experiments on our experimental testbed to validate the robustness of the proposed security scheme. The resulting accuracy with $F_s = 10$ kHz is depicted by the ROC[2] curves in Figure 8a for a range of α. From this figure it is clear that between $\tau_{attack} = 500$ and 700 μs is all that is necessary for PyCRA to accurately distinguish attacked signals from normal signals, if α is chosen appropriately. With α set to a predetermined value, we can vary F_s as shown in Figure 8b[3]. These results show that increasing F_s from 10 kHz to 30 kHz reduces required time for detection to between $\tau_{attack} = 100$ and 200 μs.

Repeating these experiments for the *advanced spoofing attack* [T3] yields similar results. In fact, there is no fundamental difference between the two in terms of attack detection; this is governed by the dynamics of the attacker's actuator rather than the nature of the attack itself.

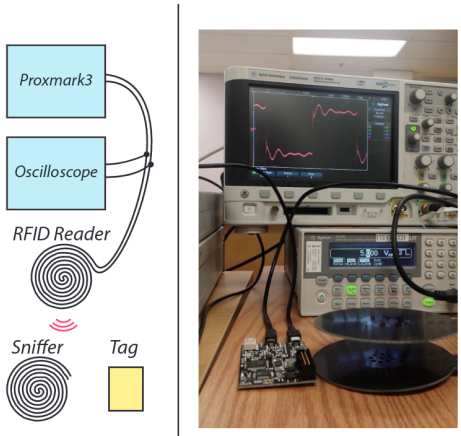
It is important to evaluate this detection accuracy (which is our security guarantee) in terms of the physical delay property τ_{attack} of the attacker model. In practice, the state-of-the-art in low-dimension, high Q-factor hardware that provide enough power to carry out a spoofing attack will have $\tau \gg 200\mu s$[4]. From Figure 8b it is apparent that PyCRA has good performance for this range of practical physical delays.

Moreover, the results we have shown thus far use a relatively low sampling frequency (high end micro controllers can operate in the range of 200 kHz). As illustrated by Figure 8b, higher sampling rates result in reduced attack detection times. However, using low sampling frequencies in our case study serves to illustrate the efficiency of the proposed detection mechanism.

[2]A Receiver Operating Characteristic (ROC) is a visual aid for evaluating the accuracy of binary classifiers in terms of both true and false positive rates.

[3]The F_1 score is a statistical measure of a binary classifier that measures the classifier accuracy in terms of precision and recall.

[4]These values were obtained by surveying a range of state-of-the-art, commercially available components.



(a) (b)

Figure 9: The schematic used in the RFID eavesdropping case study (a) and corresponding hardware setup (b). The setup contains two low frequency antennas (one for the RFID reader and one for the eavesdropper) along with a Proxmark3 board running the PyCRA detection algorithm. The analog signal is also captured by a digital oscilloscope for visualization

6. CASE STUDY (2): DETECTION OF PASSIVE EAVESDROPPING ON RFID

In this section, we discuss detection of passive eavesdropping attacks on active sensors. In this scenario, an adversary *listens* or *records* the same physical signals captured by the sensor. Indeed this type of attack satisfies assumptions A1–A3 described in Section 3 and hence it will be useful to extend PyCRA to such cases.

6.1 Passive Eavesdropping on RFID

In this section, we use radio-frequency identification (RFID) as an example where successful passive attacks can have severe consequences. RFID systems are commonly used to control access to physical places and to track the location of certain objects. An RFID system consists of two parts: a reader and a tag. The RFID reader continuously sends a magnetic signal that probes for existing tags in the near proximity. Once a tag enters the proximity of the reader, the tag starts to send its "unique" identifier to the reader by modulating the magnetic probe signal.

RFID tags can be classified as either passive or active based on the source of their energy. While passive tags rely on the energy transmitted by an RFID reader in order to power their electronics, active tags have their own power supplies. As a result, active tags can be equipped with computational platforms that run cryptographic and security protocols to mitigate cloning attacks [5]. On the other hand, passive tags do not enjoy these luxuries and therefore are more prone to cloning attacks.

Cloning of passive RFID tags can take place in one of two forms. In the first, the attacker uses a *mobile* RFID reader and attempts to place it near the RFID tag. The tag innocently responds to the communication issued by the adversarial reader and sends its unique identifier. The other form of attack carried out against RFID systems is to eavesdrop on the communication between the tag and a legitimate reader. RFID protective shields and blockers [27, 15] are commonly used as countermeasure to the first form of cloning attacks discussed above. Unfortunately, these shields are of no use

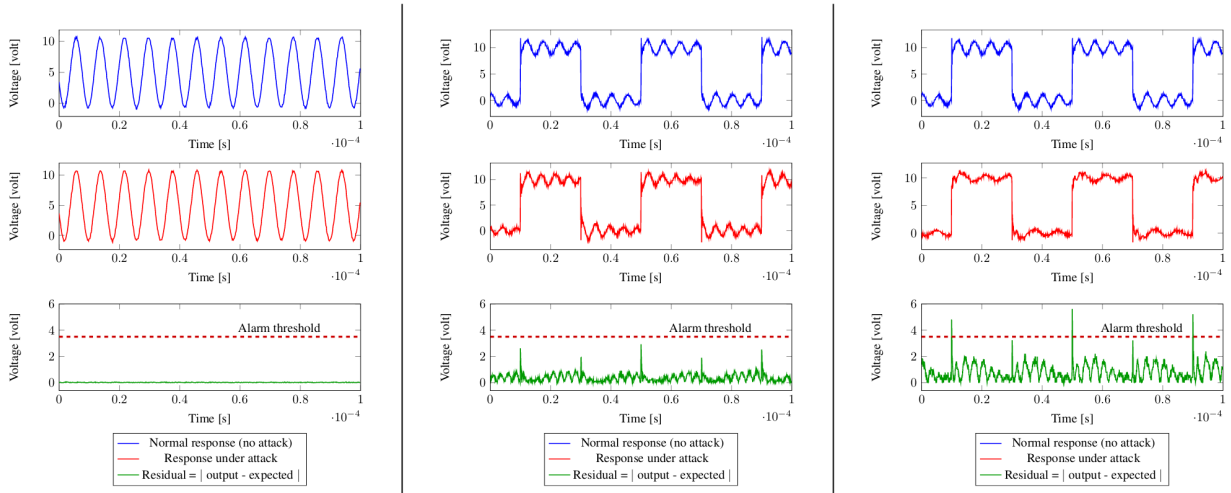


Figure 10: Results of applying PyCRA to detect the existence of an eavesdropper in the proximity of an RFID reader. (a) Results of using standard 125KHz signal for detection. (b) Results of using PyCRA when only an RFID tag is present in the proximity of the PyCRA-enabled reader, and (c) Results of using PyCRA in detecting eavesdropping when both an RFID tag and an eavesdropper antenna are present in the proximity of the PyCRA-enabled reader. Top figure shows the response to the physical challenges when no eavesdropper is placed in the proximity of the RFID reader. The middle figure shows the response to the physical challenges when (a) an eavesdropper antenna (b) passive tag only (c) passive tag + eavesdropper antenna are placed in the proximity of the reader. Finally, the bottom figure shows the value of the residuals calculated by PyCRA along with the alarm threshold.

against the second type of attacks, because the user is obliged to remove the protective shield before using the tag with the *legitimate* RFID reader, at which time an adversary can successfully eavesdrop.

To carry out an eavesdropping attack, a sniffing device must be placed in close proximity to the RFID reader so that it can measure the electromagnetic waves transmitted by the reader and reflected by the tag. In the following results, we show how PyCRA is able to detect the existence of such an attack, allowing the reader to disable communication with the tag before revealing private information.

6.2 Using PyCRA to Detect Eavesdroppers

Recall from the physics of electromagnetic waves that antennas placed within close proximity will always affect each other's electrical signals to some degree. This is a fundamental law in physics known as *magnetic coupling* [2] and is used in the design of RFID. Note that, similar to the physical delays, the magnetic coupling assumption is a fundamental limitation that the attacker cannot overcome. Hence, we can use the PyCRA detection algorithm outlined in Section 3 by computing the residual between the model (which assumes magnetic coupling only with the RFID tag) and the sensor output which suffers from magnetic coupling with the eavesdropping antenna. This is shown in the experimental results in the next subsection.

6.3 Hardware Setup and Real-time Results

Figure 9 shows the hardware setup used to carry out this case-study. In this setup, two identical low frequency RFID antennas are used. The first RFID antenna is used as the *legitimate* RFID reader while the second is used to eavesdrop. We built a PyCRA-enabled RFID reader on top of the open source RFID Proxmark3 board, adding random modulation to the probing signal and constructing the appropriate models as outlined in Section 3.

Figure 10a (top) shows the received electromagnetic wave of an RFID reader operating in the classical operational mode. In this mode, the RFID reader generates the standard 125KHz sinusoidal wave that is used to communicate with the RFID tag. Figure 10a (middle) shows the resulting electromagnetic wave when an eavesdropper uses an identical antenna to listen. In this case it is hard to distinguish between the two waves and hence hard to detect the existence of an eavesdropper. This is better illustrated by the residual between the two waves as shown by the residual in Figure 10a (bottom).

On the other hand, Figures 10b and 10c show the result of the same case study when PyCRA is used with and without an eavesdropper present, respectively. In this mode, the PyCRA algorithm occasionally halts the normal operation of the RFID reader and switches to detection mode. In this mode, PyCRA selects randomized periods of time to issue the physical challenges by switching the RFID antenna from on to off and from off to on.

In order to select an appropriate alarm threshold, we first study the effect of the magnetic coupling between the reader and the tag in the absence of an eavesdropper. This is shown in Figure 10b where the alarm threshold is chosen such that no alarm is triggered when the effect of the magnetic coupling—the residual between the "no tag" response (top) and the response with a tag (middle)—is within the acceptable range. This acceptable residual range allows for coupling induced by the tag only. Any increase on top of this allowable threshold is then attributed to the existence of an excessive magnetic coupling due to an eavesdropper.

Figure 10c (middle) shows the response to the same set of physical challenges when the attacker places an eavesdropping antenna in the proximity of the RFID reader while the tag is also present. Thanks to the physical challenge produced by PyCRA, the magnetic coupling produced by the eavesdropper antenna is now easily detected. This can be shown in Figure 10c (bottom) which shows the residuals between the expected output and the measured signal exceeding the alarm threshold.

We recorded over 1 hour of RFID measurements with varying distances of the malicious eavesdropping antenna. Of those experiments where the attacker was close enough to observe the RFID communication, PyCRA successfully detected the existence of an

attacker with 100% accuracy. Intuitively, if the attacker successfully measures the RFID signal, he has consequently removed enough energy from the channel to trigger an alarm.

7. CASE STUDY (3): EXTENDING PYCRA FOR ATTACK RESILIENCE

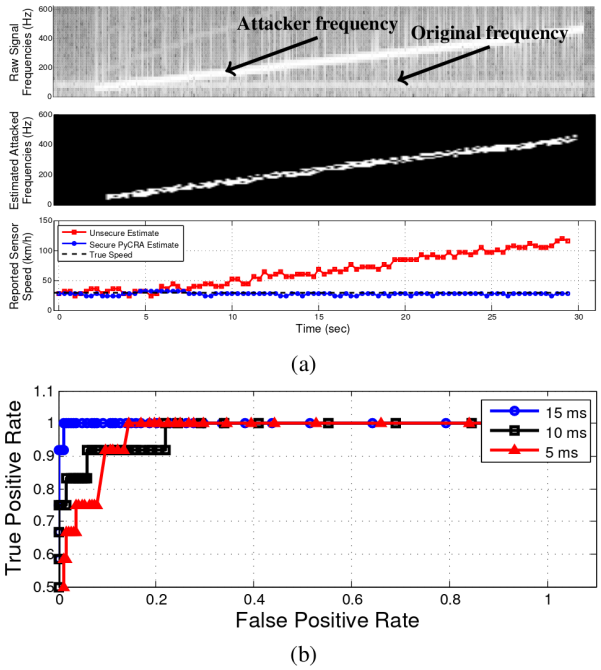
In some scenarios it is possible to extend the PyCRA authentication system to provide attack resilience as well. Here the goal is to design a sensor whose estimate of a physical phenomenon remains robust in the face of a range of physical attacks. One way to mitigate the effects of an attack is by characterizing the attack signal during the confusion phase, when the attacking signal remains detectable for τ_{attack}. This characterization depends on the properties of the sensor and signals in question, but it is in general made easier if the attack signal is sparse in some domain. One such attack resilience scheme for magnetic encoders is described below.

The signal reflected by a magnetic encoder's tone ring is a sinusoidal wave dominated by a single frequency component. If we consider the frequency domain representation of the measured signal over a window, we expect to see the energy concentrated at one frequency corresponding to that of the tone ring. However, in the existence of an attacker and using Fourier analysis, we can reasonably expect to observe energy concentrated at multiple frequencies. The sensor must be able to distinguish between the correct frequency and those of the attacker. If the attack signal itself is reasonably sparse in some domain—e.g., the Fourier domain in this example—we may obtain an accurate model of that signal over a short time period, aided by the confusion period and delay τ_{attack}. For example, we can use our earlier χ^2 estimator as an attack indicator for any given frequency component using the recursive Discrete Fourier Transform (DFT) with the following form:

$$Y_k(t+1) = e^{(j2\pi k/N)}Y_k(t) + e^{(-j2\pi k(N-1)/N)}y(t) - e^{(j2\pi k/N)}y(t-N) \quad (5)$$

where $Y_k(t) \in \mathbb{C}^N$ is the kth component of the N-point DFT of the sensor output $y(t)$ at time $t \in \mathbb{N}_0$. The χ^2 detector uses the sensor model along with (5) to predict the natural response of the tone gear (in the case of no attack) denoted $\widehat{Y}_k(t)$. We define the χ^2 residual in the frequency domain as $Z_k(t) = |Y_k(t)| - |\widehat{Y}_k(t)|$, allowing us to define thresholds to again indicate when an attack is detected and, now, the nature of that attack in terms of its its frequency components.

We implemented the scheme described above and again tested against a number of spoofing attacks in more than 90 experiments. Figure 11a illustrates the result of these experiments for a swept frequency attack, where the malicious (red) signal is accurately detected and therefore easily subtracted from the measured signal, providing a more robust estimate of the signal of interest. Figure 11b shows the accuracy of this prediction in terms of true and false positive rates as a function of τ_{attack}. Here we demonstrate accurate attack characterization and mitigation even for τ_{attack} delays as low as 5 ms. In addition, the PyCRA-secured magnetic encoder remains robust even in the face of an active cancellation spoofing attack, where the attacker attempts to use destructive interference to destroy the signal of interest. Again, during the confusion period we can actively characterize the attack signal and subtract its effects.



(a)

(b)

Figure 11: A visual tour of the PyCRA resilience scheme against the swept frequency attack (a) and attack prediction accuracies with a range of τ_{attack} values.

8. DISCUSSION

We have shown that PyCRA is able to detect and possibly mitigate different attacks on physical signals. In this section, we provide some distilled discussion thrusts by comparing PyCRA to the existing literature, revisiting the connection between physics and security guarantees and the effect of using PyCRA on the overall system performance.

8.1 Comparison With Literature

We can classify previous work on secure sensing based on the redundancy needed to provide security. For example the work described in [7, 30, 35] provides security through network and sensor redundancy. Additional research has explored fusing redundant *measurements* to provide secure localization [32, 21] and secure time synchronization [24]. The work presented in this paper falls within a second category in which no redundancy is needed to provide security. Indeed, providing security using single measurements is complementary to the security guarantees provided by redundancy based techniques. We can further classify this category into two subcategories depending on the dependency of cryptographic constructs. As described in previous sections, cryptographic constructs fail to counteract active attacks taking place in the analog domain—a gap for which PyCRA has been designed and implemented. While PyCRA is novel in bringing security against active attacks on sensory data in the analog domain, there exist similar ideas in the literature for other types of active attacks. For example, the work reported in [34] also exploits physical properties in order to provide secure localization mechanism for wireless sensor networks without relying on redundancy.

Numerous results in the literature have described and implemented counter-measures for passive attacks (e.g, eavesdropping), especially in the RFID literature. For example, in the work reported in [37, 29] the eavesdropper reveals itself through power leakage (e.g., oscillators), provided such power measurements can be obtained. In contrast, PyCRA does not assume the existence of power consumption side-channels. Rather, the experiments shown in Figure 10 demonstrate an environment with only a passive eavesdrop-

ping antenna and no such power leakage. Another example is the work reported in [11] which considers an active relay attack where the attacker actively sends information. The detection scheme is based on detecting the action of sending. As discussed above, our system works for completely passive attacks. However, PyCRA is not designed to protect against a passive eavesdropper that measures backscattered waves as discussed in [12] (compared to magnetic coupling eavesdropping). Typically RFID range is small and we make the case that PyCRA creates security for short-range operation. Indeed, some of the ideas in literature can serve to complement the contributions in PyCRA.

8.2 Physics, Randomness, & Security

The underlying ideas behind PyCRA utilize accurate mathematical models that capture the physics and dynamics of active sensors. With these models in hand, PyCRA is able to isolate the response of the attacker from the response of the environment when challenged by the proposed secure sensor. The success of the system presented in this paper lends credence to the notion of physical challenge-response authentication as a whole. Section 1 made brief mention of several additional sensing modalities where PyCRA can provide improved security. In fact, adaptation of the methods described in this work to these sensors is not difficult—one need only revisit the physical models involved in order to derive a method of probing the physical environment and measuring the environmental response.

The results presented in the previous sections demonstrate the key strengths of PyCRA—namely that it uses fundamental properties of physics (physical delays and noise) along with mathematical models to provided the following security aids (i) timely and accurate detection of external, potentially malicious sources of signal interference, (ii) resilience via accurate estimation of malicious spectra, and (iii) accurate detection of passive eavesdropping attacks. The success of these security mechanisms are bolstered by physical limitations such as delays (τ_{attack}) and change point detection, for which we have derived theoretical guarantees.

Another major factor in the security provided by PyCRA is the amount of randomness used in generating the physical challenges. The relationship between randomness and security guarantees is a classical relationship that appears in most cryptographic security protocols. However, a fundamental difference in this scheme is that PyCRA relies only on private randomness compared to shared randomness in classical cryptographic authentication schemes. This is a consequence of the "passivity" property of the measured entity exploited by PyCRA. This in turn eliminates the classical problem of random data distribution (e.g., key distribution) and thus increases the security provided by the proposed system.

8.3 Overcoming Negative Effects of PyCRA

A PyCRA-secured sensor will occasionally cease measurement of the physical phenomena it is measuring. This can negatively affect the overall sensor performance. However, measuring any physical phenomenon requires a sampling rate dictated by its bandwidth—e.g., magnetic encoders require 50Hz sampling-rate—while electronics can operate much faster (e.g., 10kHz in our experiments), and we exploit this by performing authentication using oversampling, thereby performing multiple authentications within a single physics-dictated sampling-period. Once the sensor is authenticated, we can select any of the authenticated measurements collected within the sampling-period and provide this as the 'secure' sensor measurement without affecting performance. If the sensor detects an attack, then attack-resilient estimation/computation (shown in Section 7) needs to be carried out. Again, these computations operate at a higher sampling rate. Higher sampling rate comes at the cost of increased power consumption, but this is perhaps a reasonable price to pay for security.

9. CONCLUSION

We have presented PyCRA, a physical challenge-response authentication method for active sensors. The driving concept behind PyCRA is that, through random physical stimulation and subsequent behavior analyses, we are able to determine whether or not a sensor is under attack and, in some cases, remain resilient to attacks. This work further describes ways in which the PyCRA scheme can be applied to (1) passive eavesdropping attacks, (2) simple spoofing attacks, and (3) more advanced spoofing attacks. We have demonstrated the effectiveness of PyCRA for more than 90 experiments on physical hardware focusing on three case studies: magnetic encoder attack detection, magnetic encoder attack resilience, and RFID eavesdropping detection. Our results from these case studies indicate that physical challenge-response authentication can accurately and reliably detect and mitigate malicious attacks at the analog sensor level. Finally, we believe the results discussed in this work lend credence to the notion of physical challenge-response systems in general, advocating its adoption for active sensors in general where secure operation is a critical component. More broadly, PyCRA offers security to a wide array of systems (not just sensors) where inputs are mapped to outputs by well-understood models.

Acknowledgments

This material is based upon work supported by the NSF under award CNS-1136174. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of NSF or the U.S. Government.

10. REFERENCES

[1] R. Anderson and M. Kuhn. Tamper resistance: A cautionary note. In *Proceedings of the 2Nd Conference on Proceedings of the Second USENIX Workshop on Electronic Commerce - Volume 2*, WOEC'96, pages 1–11, Berkeley, CA, USA, 1996. USENIX Association.

[2] J. Brauer. *Magnetic Actuators and Sensors*. Wiley, 2006.

[3] A. A. Cárdenas, S. Amin, and S. Sastry. Research challenges for the security of control systems. In *Proceedings of the 3rd conference on Hot topics in security*, HOTSEC'08, pages 6:1–6:6, Berkeley, CA, USA, 2008. USENIX Association.

[4] G. Dán and H. Sandberg. Stealth attacks and protection schemes for state estimators in power systems. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pages 214–219, 2010.

[5] T. Dimitriou. A lightweight rfid protocol to protect against traceability and cloning attacks. In *Security and Privacy for Emerging Areas in Communications Networks, 2005. SecureComm 2005. First International Conference on*, pages 59–66, Sept 2005.

[6] F. Dorfler, F. Pasqualetti, and F. Bullo. Distributed detection of cyber-physical attacks in power networks: A waveform relaxation approach. In *allerton*, pages 1486–1491, Allerton, IL, USA, Sept. 2011.

[7] H. Fawzi, P. Tabuada, and S. Diggavi. Secure estimation and control for cyber-physical systems under adversarial attacks. *IEEE Transactions on Automatic Control*, 59(6):1454–1467, June 2014.

[8] Ford Motor Company. Fault detection and isolation in automotive wiring harness including dedicated test line, Nov. 23 1993. US Patent 5,264,796.

[9] J. Fraden. *Handbook of Modern Sensors: Physics, Designs, and Applications (Handbook of Modern Sensors)*. SpringerVerlag, 2003.

[10] C. Grimes, E. Dickey, and M. Pishko. *Encyclopedia of Sensors*. American Scientific Publishers, 2006.

[11] S.-B. Hamida, P.-H. Thevenon, J.-B. Pierrot, O. Savry, and C. Castelluccia. Detecting relay attacks in rfid systems using physical layer characteristics. In *Wireless and Mobile Networking Conference (WMNC), 2013 6th Joint IFIP*, pages 1–8, April 2013.

[12] G. P. Hancke. Practical eavesdropping and skimming attacks on high-frequency rfid tokens. *J. Comput. Secur.*, 19(2):259–288, Apr. 2011.

[13] T. Humphreys, B. Ledvina, and M. Psiaki. Assessing the spoofing threat: Development of a portable gps civilian spoofer. In *Technical Report*. Cornel University, 2008.

[14] V. M. Igure, S. A. Laughter, and R. D. Williams. Security issues in SCADA networks. *Computers and Security*, 25(7):498 – 506, 2006.

[15] A. Juels. Rfid security and privacy: a research survey. *Selected Areas in Communications, IEEE Journal on*, 24(2):381–394, Feb 2006.

[16] T. Kim and H. Poor. Strategic protection against data injection attacks on power grids. *Smart Grid, IEEE Transactions on*, 2(2):326–333, 2011.

[17] O. Kömmerling and M. G. Kuhn. Design principles for tamper-resistant smartcard processors. In *Proceedings of the USENIX Workshop on Smartcard Technology on USENIX Workshop on Smartcard Technology*, WOST'99, pages 2–2, Berkeley, CA, USA, 1999. USENIX Association.

[18] O. Kosut, L. Jia, R. Thomas, and L. Tong. Malicious data attacks on the smart grid. *Smart Grid, IEEE Transactions on*, 2(4):645–658, 2011.

[19] D. Kune, J. Backes, S. Clark, D. Kramer, M. Reynolds, K. Fu, Y. Kim, and W. Xu. Ghost talk: Mitigating emi signal injection attacks against analog sensors. In *Security and Privacy (SP), 2013 IEEE Symposium on*, pages 145–159, May 2013.

[20] I. D. Landaue, R. Lozano, M. M'Saad, and A. Karimi. *Adaptive Control: Algorithms, Analysis and Applications*. Communications and Control Engineering. Springer, June 2011.

[21] D. Liu, P. Ning, and W. Du. Attack-resistant location estimation in sensor networks. In *Information Processing in Sensor Networks, 2005. IPSN 2005. Fourth International Symposium on*, pages 99–106, April 2005.

[22] Y. Liu, P. Ning, and M. K. Reiter. False data injection attacks against state estimation in electric power grids. In *Proceedings of the 16th ACM conference on Computer and communications security*, CCS '09, pages 21–32, New York, NY, USA, 2009. ACM.

[23] L. Ljung. *System Identification: Theory for the User*. Pearson Education, 1998.

[24] M. Manzo, T. Roosta, and S. Sastry. Time synchronization attacks in sensor networks. In *Proceedings of the 3rd ACM Workshop on Security of Ad Hoc and Sensor Networks*, SASN '05, pages 107–116, New York, NY, USA, 2005. ACM.

[25] R. Mehra and J. Peschon. An innovations approach to fault detection and diagnosis in dynamic systems. *Automatica*, 7(5):637 – 640, 1971.

[26] F. Miao, M. Pajic, and G. J. Pappas. Stochastic game approach for replay attack detection. In *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*, pages 1854–1859, Dec 2013.

[27] A. Mitrokotsa, M. Rieback, and A. Tanenbaum. Classifying rfid attacks and defenses. *Information Systems Frontiers*, 12(5):491–505, 2010.

[28] Y. Mo and B. Sinopoli. Secure control against replay attacks. In *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*, pages 911–918, Sept 2009.

[29] A. Mukherjee and A. Swindlehurst. Detecting passive eavesdroppers in the mimo wiretap channel. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pages 2809–2812, March 2012.

[30] F. Pasqualetti, F. Dorfler, and F. Bullo. Attack detection and identification in cyber-physical systems. *Automatic Control, IEEE Transactions on*, 58(11):2715–2729, Nov 2013.

[31] S. Ravi, A. Raghunathan, and S. Chakradhar. Tamper resistance mechanisms for secure embedded systems. In *VLSI Design, 2004. Proceedings. 17th International Conference on*, pages 605–611, 2004.

[32] N. Sastry, U. Shankar, and D. Wagner. Secure verification of location claims. In *Proceedings of the 2Nd ACM Workshop on Wireless Security*, WiSe '03, pages 1–10, New York, NY, USA, 2003. ACM.

[33] Y. Shoukry, P. D. Martin, P. Tabuada, and M. B. Srivastava. Non-invasive spoofing attacks for anti-lock braking systems. In *Workshop on Cryptographic Hardware and Embedded Systems 2013*, G. Bertoni and J.-S. Coron (Eds.): CHES 2013, LNCS 8086, pages 55–72. International Association for Cryptologic Research, 2013.

[34] D. Singelee and B. Preneel. Location verification using secure distance bounding protocols. In *Mobile Adhoc and Sensor Systems Conference, 2005. IEEE International Conference on*, pages 7 pp.–840, Nov 2005.

[35] K. Sou, H. Sandberg, and K. Johansson. On the exact solution to a smart grid cyber-security analysis problem. *Smart Grid, IEEE Transactions on*, 4(2):856–865, 2013.

[36] A. G. Tartakovsky and V. V. Veeravalli. General asymptotic bayesian theory of quickest change detection. *Theory Probab. Appl.*, 49(3):458 –497, 2005.

[37] B. Wild and K. Ramchandran. Detecting primary receivers for cognitive radio applications. In *New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005. 2005 First IEEE International Symposium on*, pages 124–130, Nov 2005.

[38] A. S. Willsky. A survey of design methods for failure detection in dynamic systems. *Automatica*, 12(6):601 – 611, 1976.