IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 2 November 2022, accepted 11 December 2022, date of publication 28 December 2022, date of current version 18 January 2023.

Digital Object Identifier 10.1109/ACCESS.2022.3233010

RESEARCH ARTICLE

Synthesizing Pareto-Optimal Signal-Injection Attacks on ICDs

VEENA KRISH[1], NICOLA PAOLETTI[2], SCOTT A. SMOLKA[1], AND AMIR RAHMATI[1]
[1]Stony Brook University, Stony Brook, NY 11794, USA
[2]Department of Informatics, King's College London, WC2B 4BG London, U.K.

Corresponding author: Veena Krish (kveena@cs.stonybrook.edu)

This work was supported in part by the Air Force Research Laboratory under Grant FA9550-22-1-0029 and Grant FA9550-22-1-0450; in part by NSF under Award ITE-2134840, Award CCF-1918225, and Award CPS-1446832; and in part by EPSRC under Award EP/W014785/2.

ABSTRACT Implantable Cardioverter Defibrillators (ICDs) are medical cyber-physical systems that monitor cardiac activity and administer therapy shocks in response to sensed irregular electrograms (EGMs) to prevent cardiac arrest. Prior work has shown that the analog sensors used in these systems are vulnerable to *signal-injection attacks*. Such attacks induce morphological changes in EGM measurements that disrupt the normal behavior of the ICD's control software and cause the device to administer incorrect therapy. Existing work has primarily focused on the feasibility of such attacks and has not examined how they can be systematically devised. In this paper, we introduce *InjectICD*, a model-based framework for the systematic construction of stealthy signal-injection attacks that can thwart ICD control software. InjectICD solves the problem of synthesizing attack signals as one of multi-objective optimization, thereby allowing it to identify Pareto-optimal signal-injection templates that maximize the probability of attack success while simultaneously applying minimal modifications to the original EGM. We evaluate InjectICD on an ICD algorithm currently implemented in devices from a major ICD manufacturer. We show that InjectICD can construct such attack templates for various heart conditions and under different adversary capabilities, while also demonstrating that our approach generalizes to unseen EGM signals. Our results highlight the urgent need for ICD manufacturers to incorporate defenses against signal-injection attacks.

INDEX TERMS Medical device security, signal-injection attacks, Pareto-optimal attacks.

I. INTRODUCTION

Medical devices have experienced rapid advancement over the last four decades, evolving from simple analog devices into complex cyber-physical systems that continuously monitor biological signals to diagnose and administer therapy With over 200,000 implanted every year [1], *Implantable Cardioverter Defibrillators* (ICDs) are a prime example. ICDs sense *intracardiac electrograms* (EGMs), run embedded software to detect treatable arrhythmia, and deliver appropriate therapy in the form of an electrical shock to prevent sudden cardiac arrest. ICDs generally use a software-based *discrimination algorithm* to decide whether therapy should be delivered based on physiological criteria for the classification of arrhythmias. These algorithms were designed to reproduce physician diagnoses and do not consider the risk posed by an adversary intent on manipulating their decision processes.

The ever-increasing sophistication of ICDs has enabled them to treat a wide array of cardiac arrhythmogenic disorders. Their inherent complexity has also made them prone to security and privacy attacks. Halperin et al. [2] demonstrated that ICDs can be reprogrammed by unauthorized users using commercial software radios. More recently, researchers managed to gain control of a pacemaker/ICD by exploiting vulnerabilities in the device's remote monitoring infrastructure, resulting in the recall of half a million cardiac devices by the FDA [3]. Kune et al. [4] showed that ICDs are vulnerable to signal-injection attacks, where electromagnetic interference signals are used to inhibit pacing and induce defibrillation. While this research provided an important proof-of-concept for ICD attacks, it did not evaluate the expected success rate

The associate editor coordinating the review of this manuscript and approving it for publication was Wenjie Feng.

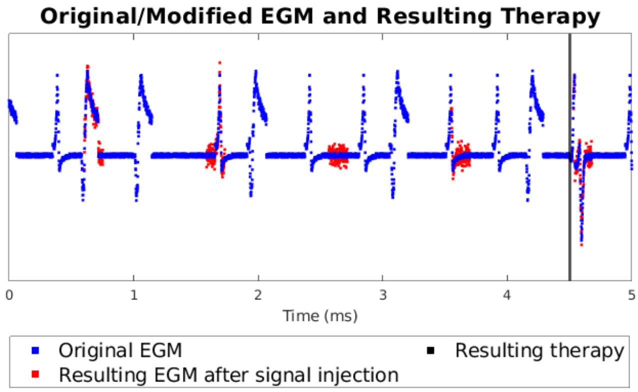

FIGURE 1. Example signal-injection and resulting modified signal for the ventricular lead. The added pulses of Gaussian noise (in red) lead to an inappropriate therapy shock (black vertical line).

of attacks under various realistic threat models, nor did it take into consideration the *stealthiness* of injected signals.

In this paper, we present *InjectICD*, a framework for the systematic derivation of signal-injection attacks on EGM signals designed to cause incorrect ICD therapy decisions. Unlike previous work that focused on demonstrating the feasibility of such attacks, InjectICD is concerned with the synthesis of signal-injection attacks that are at once *effective* (introduce inappropriate therapy), *stealthy* (involve minimal changes to the true EGM, hence minimizing the likelihood of detection), and *realistic* (can be carried out under a reasonable threat model). Figure 1 presents an example of an attack generated by InjectICD, where an unnecessary therapy shock is administered in response to four brief pulses of injected noise, shown in red.

InjectICD formulates the search for an optimal attack as a multi-objective optimization problem, maximizing both effectiveness and stealthiness objectives. We represent signal-injection attacks using periodic signal templates optimized with respect to their characteristic parameters (e.g., a pulsed square wave defined by pulse width, amplitude, and frequency of pulses). To systematically determine such attack parameters, InjectICD leverages an executable model of the target ICD algorithm evaluated on a set of training EGM signals.

Our work is inspired by that of Paoletti et al. [5], who presented a formal synthesis method for finding Pareto-optimal ICD reprogramming attacks (i.e., targeting the device-programmable parameters). We focus in this paper on a different attack space: the vulnerability of ICDs to small adversarial deviations in the input EGM.

Figure 2 provides an overview of the InjectICD framework. We model adversaries that can exploit their knowledge of the victim's pathology and of the ICD algorithm to craft optimal and patient-specific attack signals. We explore the problem under a variety of threat models: (1) *Insider attacks*, where the adversary knows a patient's heart condition and has access to their historical EGM tracings. (2) A *Knowledgeable attacker* knows the patient's heart condition but only has access to EGM tracings from a patient population that share the same heart condition as the victim. (3) *Universal attacks* can induce misclassifications without knowledge of the patient's heart condition. These three threat models are highlighted in Figure 2.

We evaluate our InjectICD framework on an ICD discrimination algorithm used by a top medical device manufacturer under 19 heart conditions. InjectICD is able to find attack parameters that are up to 71% successful on average in causing the device to deliver incorrect therapy for a given condition.

By showing how to generate stealthy and effective signal-injection attacks, our work confirms the urgency surrounding the lack of protection against spoofing attacks in current ICD devices, and highlights the need to harden these devices and their algorithms against signal injection. These attacks do not necessarily imply that the underlying algorithm is incorrect, but illustrate that ICD therapy can be significantly and dangerously impacted by adversarial inputs.

In summary, our main contributions are as follows:

- We introduce InjectICD, a novel framework for the automatic synthesis of signal-injection attacks on ICDs. InjectICD uses multi-objective optimization to derive attacks that are both stealthy and effective.
- We present three realistic threat models under which such attacks can occur and explore how attackers with various levels of knowledge can generate targeted signal-injection attacks for multiple heart conditions.
- We evaluate our framework on an ICD discrimination algorithm used by a major medical device manufacturer under 19 unique heart conditions. Our results show that InjectICD can successfully devise injection attacks tailored to specific cardiac arrhythmias, and remains effective when no prior information about the victim's condition is available.

II. BACKGROUND

ICDs are implanted medical devices that administer an electrical shock when a potentially life-threatening cardiac arrhythmia is detected. Figure 3 presents a picture of an ICD and its lead placements. These devices are implanted under the pectoral muscles in the chest and sense the heart's electrical activity via multiple leads placed within the cardiac chambers. The tracing of electrical potentials recorded by these leads is referred to as an *intracardiac electrogram* (EGM). A dual-chamber ICD (the most common kind) records three types of EGM waveforms: (1) *atrial signals* that describe the activity of the right atrium; (2) *ventricular signals* that describe the activity of the right ventricle; and a (3) *far-field shock lead* that provides a global view of the electrical activity.

ICDs rely on features of these sensed waveforms to detect tachycardia episodes (many types of arrhythmias can cause tachycardia) and administer the appropriate therapy. A number of ICDs appeal to a tree-structured decision algorithm that synthesizes medical judgments (i.e., how EGM features

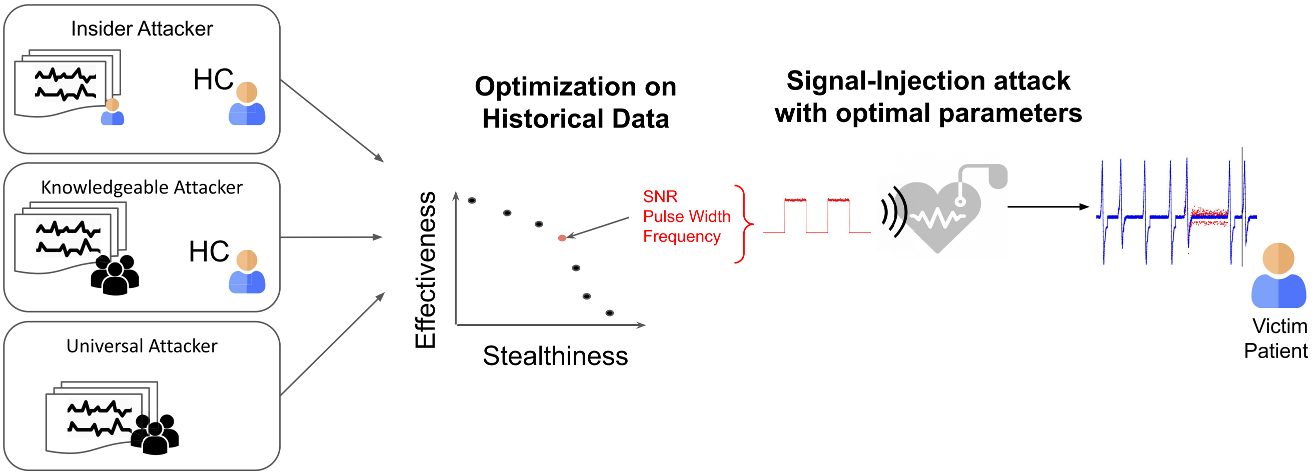

FIGURE 2. Overview of the InjectICD framework. The three threat models capture various adversarial capabilities. ***Insider Attacker*** has access to the patient's heart condition (HC) and historical data; ***Knowledgeable Attacker*** has access to the patient's HC and historical population data; ***Universal Attacker*** only has access to historical population data. Historical data are used within a multi-objective optimization framework to determine Pareto-optimal parameters (e.g., pulse amplitude, duration, frequency) that defines the signal-injection attack directed at the victim's ICD.

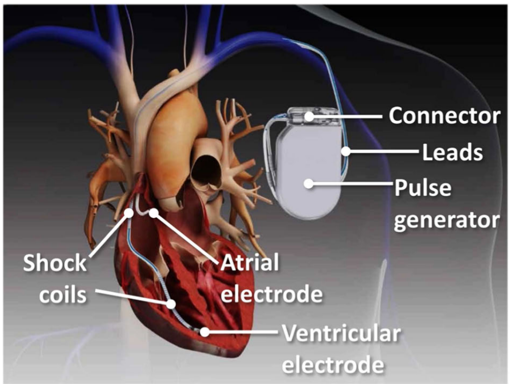

FIGURE 3. Illustration of a dual-chamber ICD and placement of atrial, ventricular, and far-field (shock) leads [5].

compare to physician-determined parameters) as various *discriminators*. These algorithms run on embedded software and rely on signal-processing methods (filtering, peak detection) for conveying morphological information.

A. SIGNAL-INJECTION ATTACKS

Analog sensors that are sensitive to electromagnetic interference expose a potential vulnerability, allowing adversaries to manipulate sensed values [6]. Cardiac devices, including ICDs, measure the heart's electrical activity using analog sensors. These signals are then typically passed through analog low-pass filters, amplified, and digitized.

This work considers an adversary that has the ability to manipulate sensor readings. In particular, we consider the case of signal-injection attacks that aim to alter the sensed readings of cardiac electrical activity by taking advantage of unshielded leads. The feasibility of such signal-injection attacks on similar medical devices has been well-studied [2], [4], [7]. These attacks can appear as baseband electromagnetic interference to avoid bandpass filtering or amplitude-modulated attacks that take advantage of systems that respond to low-power, high-frequency signals. Kune et al. [4] showed examples of signal spoofing that could inhibit pacing or introduce defibrillation shocks. Here, we are further interested in controlling the magnitude of signal injections and systematically generating malicious perturbations in such a way that they are imperceptible while affecting the decision making of the ICD discrimination algorithm.

B. EFFECT OF THERAPY DISRUPTION

In delivering therapy, an ICD can make two types of incorrect therapy decisions: *false positives*, when unnecessary therapy is administered, and *false negatives*, when the ICD fails to deliver a necessary therapy. These errors can occur in non-adversarial settings for reasons such as dislodged leads [8] or poorly calibrated parameters [9].

False positives (FPs) can lead to significant harm. Inappropriate shocks cause not just discomfort but are also associated with increased morbidity and long-term risk of death [10]. Repeated unnecessary shocks have been shown to adversely affect mental well-being, reduce the quality of life, damage cardiac tissue, and even result in proarrhythmia (provocation of a new arrhythmia) [11], [12]. Patients interviewed in a quality-of-life survey described the feeling of a shock as "an earthquake" and "being hit by a truck" [13]. Moreover, shocks significantly deplete the device's battery, thus reducing the ICD's lifetime and causing early re-implantation.

False negatives (FNs) can have even more severe consequences. Without defibrillation, ventricular fibrillation leads to sudden cardiac death. Delayed therapy is linked to the degradation of patient outcomes,with therapy delayed for more than two minutes associated with a significantly lower probability of survival after 24 hours [14].

III. ICD THREAT MODELS

We consider three ICD threat models spanning different attacker capabilities. See Figure 2. Each model assumes a different level of access to the target patient's medical data. All threat models assume the following:

- *Device knowledge:* The adversary has knowledge of the device model and the ability to implement a faithful (and executable) replica of its discrimination algorithm using publicly-available device manuals, reverse engineering, or corporate espionage.
- *Signal-injection capability:* The attacker can inject arbitrarily malicious signals to alter the EGM readings sensed by the ICD. We note that the signal source is always at some distance from the target ICD. Physical transmission of signals can be accounted for through the use of radio propagation models [15].
- *Training signals:* The attacker has access to a dataset of EGM signals that are used to craft signal-injection attacks. Depending on the threat model, these signals may come from the specific victim or from population-level data, such as the NIH-funded PhysioNet repository [16]. PhysioNet provides public access to datasets of cardiac traces for research purposes.

Two of our three threat models assume knowledge of the victim patient's heart condition. This information is useful, as the adversary, based on their knowledge of the discrimination algorithm, can identify features of the EGMs that are critical to making patient-specific therapy decisions. For example, the appropriate therapy classification for a patient who suffers from supraventricular tachycardia relies on a different set of discriminators than those for a patient with atrial fibrillation. As we will see, the attacker can use this knowledge to train highly patient-specific attacks by selecting EGM signals that are consistent with the victim's condition.

The measure of attack effectiveness optimized at training time can be seen as an estimate of the true effectiveness of the attack deployed in the field. The accuracy of this estimate directly depends on the amount of knowledge the attacker possesses about the victim's underlying condition. We consider the following three threat models, presented in decreasing order of the knowledge level of the victim's condition.

A **Insider Attacker – access to patient's heart condition and historical EGMs.** The Insider Attacker threat model is the most permissive. The adversary has access to the targeted patient's historical EGM data and therefore highly specific training EGMs. The adversary also knows the patient's heart condition, which can be used to tailor the attack to the victim. This threat model is applicable when the attacker has access to the patient's medical records. As many healthcare data breaches have shown recently, such a level of access is not far-fetched. As one example, personal medical data was stolen from the health insurance company Anthem Inc in 2015; the breach affected 78.8 million people [17].

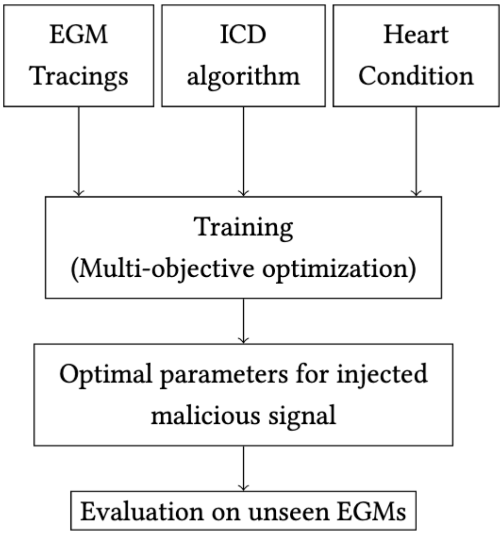
EGM Tracings
ICD algorithm
Heart Condition
Training (Multi-objective optimization)
Optimal parameters for injected malicious signal
Evaluation on unseen EGMs

FIGURE 4. Overview of the method to construct signal-injection attacks in InjectICD.

B **Knowledgeable Attacker – access to patient's heart condition and surrogate patient EGMs.** This adversary does not have access to the targeted patient's historical data. They know, however, the target's heart condition and have access to EGM data from a population of individuals with the same condition. The adversary might obtain this data from the aforementioned public repositories; for example, PhysioNet offers real EGM traces corresponding to various ventricular arrhythmias [16]. As we will show, this information is often sufficient to devise an effective attack.

C **Universal Attacker – agnostic to heart condition, access to population of EGMs.** In this threat model, the attacker deploys two types of universal attacks (i.e., agnostic to the patient-specific heart condition) that can reliably induce false-positive (*Universal Attacker-FP*) and false-negative (*Universal Attacker-FN*) misclassifications, respectively. To do so, the attacker exploits training EGMs from a population of EGM data spanning multiple heart conditions, and develops FN attacks for signals that require shock therapy, and FP attacks for those that do not. A Universal Attacker may not be able to tune the objective for attack optimization as finely as in the other two models, but they can take advantage of their knowledge of the discrimination algorithm to mount a successful attack.

1) ATTACK DETECTION MECHANISMS

An ICD keeps only a few representative signals in its storage, typically corresponding to episodes of tachyarrhythmia and delivered therapies. Thus, there are two detection mechanisms that our attacks must evade: a clinician inspecting stored waveforms during routine check-ups, or a security expert called to examine the device after a particularly suspicious adverse event. In both cases, since the device stores

EGM signals that are already perturbed by the attack (if any), and the attack leaves no other trace in the ICD, the best strategy to evade both kinds of detection is to make the perturbations imperceptible so that the stored tracings will appear benign. Hence, as we explain in the next section, it is sensible to quantify stealthiness using measures of signal distance between the original and perturbed EGMs.

IV. ATTACK DESIGN

We formulate the problem of finding stealthy and effective signal modifications as a multi-objective optimization problem designed to maximize both attack success and stealthiness. The training procedure is illustrated in Figure 4. At time t, we consider an adversary that has control over an injected signal $m(t)$, targeting an underlying signal $s(t)$, and leading to sensor reading $s'(t) = s(t) + m(t)$. The malicious signal $m(t)$ is not designed to dominate $s(t)$ but rather to cause small changes to $s(t)$ that result in incorrect therapy. In our model, the injected signal affects each of the three leads in the same way because the leads are close together (within a few centimeters); see Figure 3.

We consider time-bounded signals s, defined as functions $s : \mathbb{T} \to \mathbb{R}^n$, where $\mathbb{T} = \{0, 1, \ldots, T\} \subset \mathbb{Z}^{\geq 0}$ is the discrete-time domain, for some bound $0 < T < \infty$. We denote with $T(s)$ the length of s. The attack-synthesis problem is represented as one of multi-objective optimization, where the effectiveness of the attack is at odds with the stealthiness of the signal perturbations. Our threat models assume the availability of a set $\mathcal{D}$ of training signals, used to find the perturbation m that maximize average effectiveness and stealthiness over $\mathcal{D}$. Formally, this amounts to solving the problem:

$$\max_{m} \left[\frac{1}{|\mathcal{D}|} \sum_{s \in \mathcal{D}} f_E\left(s, s'\right), \frac{1}{|\mathcal{D}|} \sum_{s \in \mathcal{D}} f_S\left(s, s'\right) \right],$$
$$s'(t) = s(t) + m(t) \text{ for all } t \in \mathbb{T} \qquad (1)$$

where s is the training EGM sample, $f_E\left(s, s'\right)$ is the effectiveness of the spoofed EGM s' with reference to the therapy of the nominal/training EGM s, and $f_S\left(s, s'\right)$ quantifies stealthiness in terms of the distance between the nominal EGM and the spoofed one.

The distribution of training signals naturally depends on the specific threat model. In particular, insider attacks are trained using signals from the same patient used for testing. Knowledgeable attacks are trained with signals coming from different patients but with the same heart condition as the one used for testing. Universal attacks are trained by aggregating samples across multiple heart conditions that are suitable for either false positive or false negative attacks. See Table 1 for the list of heart conditions considered in our evaluation.

To facilitate the search for an optimal perturbation m, instead of searching in the space of possible signals, we consider a family of parametric signals determined by a small set of parameters a. In particular, we focus on periodic signals that can be described by their frequency, amplitude, and pulse width. This parameterization provides us with a way to explore a tractable space of possible signal-injection attacks.

For a fixed choice of parameters a, we consider multiple signals m generated using random initial offsets and define f_E and f_S in Eq. (1) as the average effectiveness and stealthiness over such generated signals.

A strength of our framework is that it provides multiple Pareto-optimal solutions of Eq. (1). In this way, the adversary can choose among the full range of attacks yielding optimal effectiveness-stealthiness trade-off. For the results presented in Section V, we assume that the attacker chooses the Pareto-optimal solution with the shortest ℓ_2-distance from the ideal (but unrealizable) attack, i.e., the one that attains both 100% effectiveness and 100% stealthiness; see Figure 7. Other post-hoc decision criteria are possible, however.

2) STEALTHINESS OBJECTIVE

We define the stealthiness of an attack in terms of the difference between the original and modified signals. Specifically, we looked at root mean squared error (RMSE), cross-correlation, and Signal-to-Interference ratio. (The last one was used by Kune et al. [4] to quantify the disturbance of EMI attacks.) We found that RMSE performed best in the search for optimal attack parameters, and therefore we present results with RMSE as our stealthiness objective.

Given that our signals are 3-dimensional, as the ICD senses ventricular, atrial and far-field EGMs (see Section II), our RMSE-based stealthiness function is defined as follows:

$$f_{RMSE}\left(s, s'\right) = \sum_{i=1}^{3} \sqrt{\frac{1}{T(s)} \sum_{t=1}^{T(s)} \left(s_i(t) - s_i'(t)\right)^2}, \qquad (2)$$

Since stealthiness (i.e., signal similarity) is antipodal to signal distance, we set $f_S\left(s, s'\right) = -f_{RMSE}\left(s, s'\right)$ in the optimization problem (1).

Note that in Figures 7 and 8, we report stealthiness as a value in the range [0, 1] defined as $1 - \bar{f}_{RMSE}$. Here, $\bar{f}_{RMSE}$ is the result of applying a min-max scaling transformation to f_{RMSE} using the value range that the RMSE attains over the entire training set during optimization.

3) EFFECTIVENESS OBJECTIVE

An attack m is effective on a signal s if the resulting perturbed signal s' induces a different therapy decision. Let $Th(s, t)$ be the output of the ICD discrimination algorithm at time t on signal s, where $Th(s, t) =$ true if the algorithm decides to deliver therapy at time t, $Th(s, t) =$ false otherwise. Then, the effectiveness objective $f_E(s, s')$ of Eq. (1) is defined as:

$$f_E(s, s') = \begin{cases} 1 & \forall t.\ Th(s,t)=\text{False and } \exists t.\ Th(s',t)=\text{True} \\ 1 & \exists t.\ Th(s,t)=\text{True and } Th(s',t)=\text{False} \\ 0 & \text{otherwise} \end{cases} \qquad (3)$$

Eq. (3) represents the idea that an effective attack signal s' is one that introduces at least one unnecessary therapy in an FP attack (top case), or suppresses at least one required therapy in an FN attack (second case).[1] Hence, the mean f_E value over the training signal used in problem (1) can be interpreted as the probability of attack success.

We also investigated alternative definitions for f_E, including "surrogate" measures that do not directly quantify effectiveness but target EGM features used in specific discriminators. For example, an FP attack can be induced by maximizing the average ventricular period—the surrogate measure in this example—thus increasing the likelihood that the ICD detects a tachyarrhythmia episode and consequently delivers unnecessary therapy. Nevertheless, the effectiveness measure given by Eq. (3) proved to be both more straightforward to calculate than alternative surrogate measures and empirically worked better for attack optimization.

V. EVALUATION

A. EXPERIMENTAL SETUP

1) DISCRIMINATION ALGORITHM

Many ICDs administer therapy based on the outcome of a *discrimination algorithm*: essentially a decision tree that operates over features extracted from three EGM signals (atrial, ventricular, and far-field leads). We choose to evaluate our InjectICD methodology on an implementation of the ICD discrimination algorithm used by a major medical device manufacturer [18], [19].

To determine whether a patient's rhythm should be treated, this algorithm uses atrial and ventricular interval timing (i.e., the time between two consecutive peaks in the atrial and ventricular EGMs, resp.), along with ventricular interval correlation analysis, to determine whether a patient's rhythm should be treated. Interval timing for each lead is based on a standard peak detection algorithm that uses bandpass filtering and thresholding.

The features used in this algorithm are extracted and stored for each ventricular interval (i.e., heartbeat). The decision to administer shock therapy is made independently at each beat interval, based on the set of features extracted from the previous 10 beats. Note that this makes the decision at the i-th beat dependent on the previous 20 beats, as the features computed at the $(i - 10)$-th beat were based on its previous 10 beats.

The ICD discrimination algorithm is presented in Figure 5. The discriminators used in the algorithm are described below. While other manufacturers may use windows of different lengths and slightly different logic, they all generally consider similar features.

In the following, VF, VT and SVT stand for ventricular fibrillation, ventricular tachycardia, and supraventricular tachycardia, respectively.

D1: Compares the ventricular period to a threshold to detect the onset of VF.

D2: Checks if a VF episode is persistent.

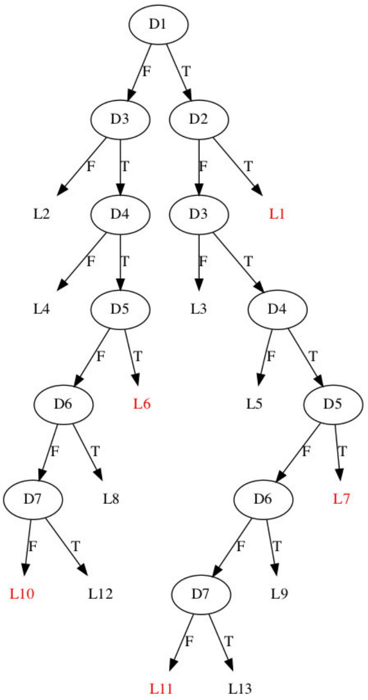

FIGURE 5. ICD discrimination tree. Discriminators D1-D7 compare features extracted from EGMs with thresholds programmed by physicians. At each ventricular interval (i.e., beat), the algorithm considers the last 10 beats to determine a path through the tree, ending at one of the leaves L1-L13. Leaves that result in a therapy shock decision are shown in red.

D3: Compares the ventricular period to a threshold to detect the onset of VT.

D4: Checks if a VT episode is persistent.

D5: Compares atrial and ventricular periods to infer the origin of the tachycardia.

D6: Discriminates between VT and SVT by comparing the morphology of the far-field shock EGM with a pre-computed normal sinus rhythm template.

D7: Discriminates between VT and SVT by analyzing atrial periods and the variance of ventricular periods.

2) SYNTHETIC EGMs

We evaluated the InjectICD framework on synthetic EGM traces generated through the method of Jiang et al. [20]. This approach has been used in several prior papers on model-based analysis of ICD therapy, including the ICD reprogramming attacks by Paoletti et al. [5]. It uses a Timed Automata model of the electrical conduction system of the heart to generate the timings of cardiac events (i.e., peaks in the EGMs), and samples the morphology of the peaks from a database of real patient records [21].

We generate EGMs for 19 distinct arrhythmias; see Table 1. We chose these heart conditions because they were previously used to evaluate the accuracy of the discrimination algorithm in a clinical trial [22]. They thus represent a realistic collection of arrhythmias requiring ICD therapy.

The 19 heart conditions captured by these signals represent a variety of arrhythmias. Some are treated by

[1] Note that only one of the two attacks applies to any EGM s: the FP condition on s ($\forall t.\ Th(s, t) =$ False) and the FN condition ($\exists t.\ Th(s, t) =$ True) are mutually exclusive.

TABLE 1. Heart conditions with general attack types.

Attack Type	Heart Condition
False Positive	Atrial Fibrilation (AF) Atrial Flutter (AFT) Premature Atrial Contractions (PAC) Supraventricular Tachycardia (SVT) Atrial Fibrilation with Premature Ventricular Contractions (AF-PVC) SVT with Wenckebach block (SVT-Wenckebach) SVT with PVC (SVT-PVC) SVT with Wenckebach block and PVC (SVT-Wenckebach-PVC) Sick Sinus Syndrome (SickSinus)
False Negative	Ventricular Tachycardia (VT) Ventricular Fibrilation (VF) Nonsustained Ventricular Tachycardia (NSVT) Atrial Fibrilation with Ventricular Fibrilation (AF-VF) Atrial Fibrilation with Ventricular Tachycardia (AF-VT) Atrial Fibrilation - Ashman complexes (AF-ashman) Ventricular Tachycardia with Atrial Flutter (VT-AFT) Premature Atrial Contractions with Ventricular Fibrilation (PAC-VF) Ventricular Tachycardia-retrograde (VT-retrograde)

low-energy anti-tachycardia pacing therapy, while others (such as VF) are treated by a single high-energy defibrillation shock. While a patient who is diagnosed with VF can still suffer from inappropriate shocks, missing one of the high-energy shocks can have more dire consequences given their condition. In Table 1, we therefore classify these 19 conditions into two groups: *False Positive* (inappropriate shocks) and *False Negative* (missing shock) attacks.

For each heart condition, we generated 25 60-second EGM traces. Each trace includes an additional 10-second prefix characterized by normal sinus rhythm, which is required for discriminator D6. Of these 25 EGMs, 5 were used for training and 20 for testing. Each training EGM was split into non-overlapping segments of 20 beats each. The resulting segments constitute our training dataset $\mathcal{D}$.

3) PARAMETERIZED ATTACK TEMPLATE SIGNALS

We considered the following two templates for periodic attack signals:

- **Square pulses** with parameters: amplitude, frequency, pulse width.
- **White Gaussian noise** with parameters: signal-to-noise ratio (SNR), frequency, pulse width.

The search space for the attack parameters was defined by the following ranges: frequency, $[1/600, 1/100]$ ms^{-1}; pulse width, $[0, 200]$ ms; SNR, $[10, 15]$ dBW (white Gaussian noise only); amplitude, $[0, 0.25]$ in normalized units. These ranges were experimentally determined by evaluating attacks on all of the test data. Attacks with an SNR above 15 (i.e., with very low noise levels) rarely succeeded, while square amplitudes above 0.25 (inducing very large perturbations) always succeeded. Ranges for the pulse widths and frequencies were chosen following similar reasoning.

4) TRAINING ALGORITHM

To solve Eq. (1), we used the NSGA-II genetic algorithm (GA) for multi-objective optimization [23] implemented in MATLAB's `gamultiobj` function. Our problem is characterized by a relatively small number of parameters but involves a challenging objective function (effectiveness f_E), which is non-convex, non-linear, and non-differentiable.[2] This kind of problem structure is well-suited to a stochastic search algorithm like NSGA-II, rather than gradient-based or constraint-solving methods. For NSGA-II, we selected a population size of 50 at each of 25 generations (with a maximum stall of 5 generations), after some hyper-parameter tuning.

Each run was performed on training samples of 20 beats and tested on 60-second samples from a withheld set. This test set was shared across all threat models (analogous to the "Victim Patient" in Figure 2), even though each threat model has different rules for attack evaluation; i.e., it uses a different distribution of training signals, as explained in Section IV.

B. EXPERIMENTAL RESULTS

1) FP AND FN ATTACKS

In Figure 6, we show examples of two successful signal-injection attacks, one for the FP case and one for the FN case. In both cases, the injected signal was generated from the solutions (amplitude, frequency, and pulse-width parameters) for a pulsed square wave yielded by our multi-objective optimization formulation. The injected signals are displayed at the top of the figure. The ventricular and atrial components of the original EGM are shown in blue, overlaid by corresponding components of the perturbed signal (red). Discriminator values are also shown for the original and modified signals for each example: at each sensed ventricular beat, the discrimination algorithm evaluates D1-D7 as described in Section V-A1.

The differences in these values explain the FP case (inappropriate therapy): additional VF episodes are sensed, as the injected wave induces changes to the legitimate signal that cause the calculated ventricular periods to shorten (i.e., increased heart rate). Discriminators D1-D4, which detect the onset and persistence of a VF episode, reflect the change. This tricks the discriminator into detecting a persistent VF episode and consequently (and incorrectly) delivering therapy.

The FN case is arguably more subtle: the attack induces changes to the signal by increasing the ventricular period length by a small amount at each of the last 5 intervals (i.e., beats are detected slightly after). These changes caused the VT duration check at D4 to fail, suppressing the therapy that would have been applied during a VT episode. This shortened VT interval length was not sustained in the modified signal.

2) ATTACK SYNTHESIS AND SELECTION

InjectICD was successful in synthesizing attacks for all threat models, and these attacks were found to generalize well to unseen data. Each training run yielded a set of Pareto-optimal solutions, where a solution corresponds to a particular set of values of the signal-injection template parameters. An example of a Pareto front is displayed in Figure 7. The plot clearly

[2] This is not just because the definition of effectiveness (Eq. (3)) depends on a logical predicate, but also because the ICD output itself $Th(s, t)$ depends on the execution of the discrimination tree, making it non-differentiable.

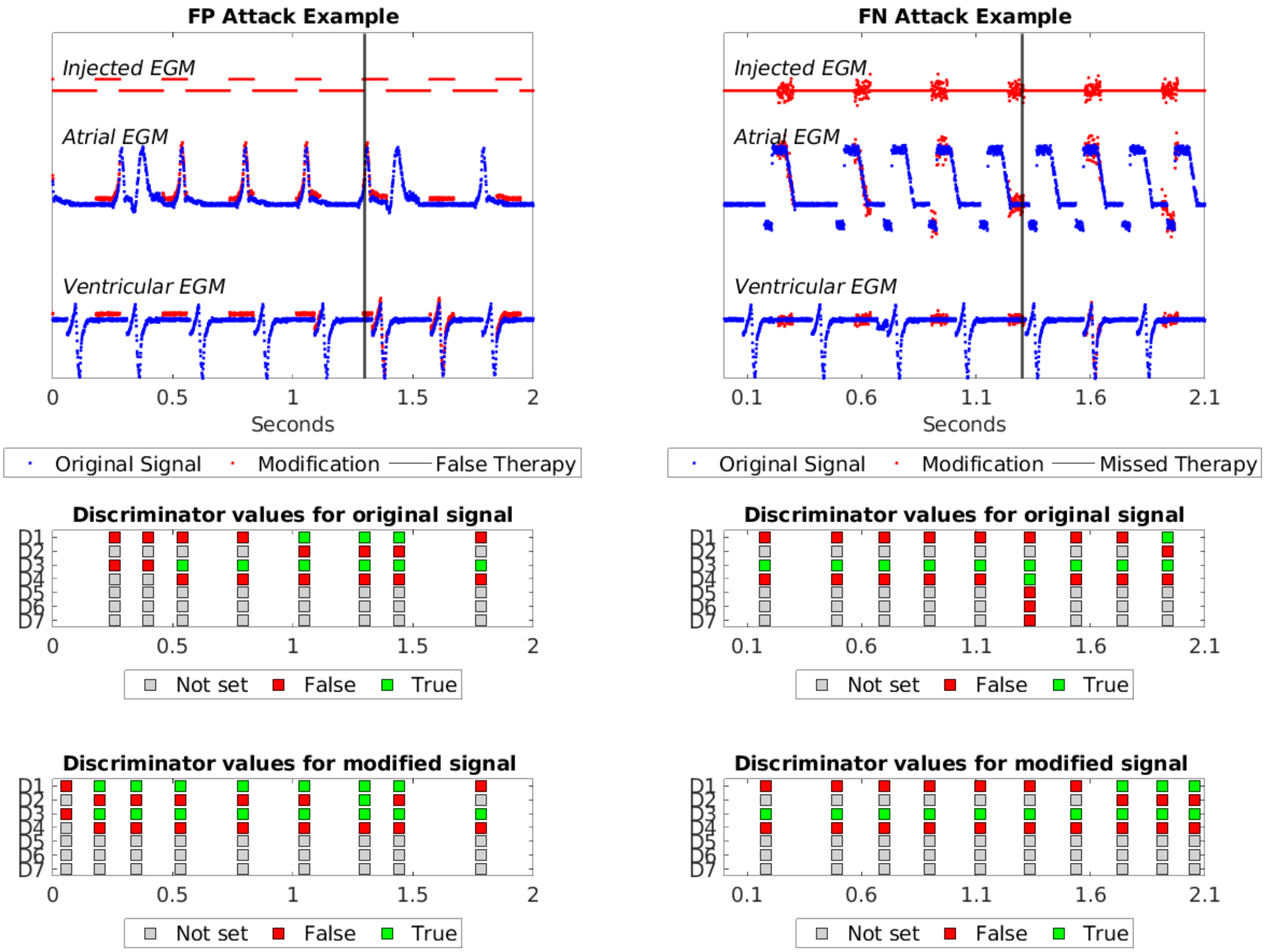

FIGURE 6. Examples of an FP attack (left) and an FN attack (right). The Discriminator-value heatmaps below the EGMs are reflective of the values of discriminators 1-7 (true values in green, false values in red), and illustrate the changes induced by the signal injection. In the FP attack, the injected square pulse causes additional VF episodes to be detected, resulting in an inappropriate shock around 1.3 seconds (6th beat). The first table shows the values of discriminators for the original EGM signal: the first four D1 values are false. The second table shows the values of discriminators after the attack: D1 and D2 are switched to True. In the FN attack, the pulse-injected signal elongated the ventricular periods in the EGM window shown. Originally, a VT episode would have persisted after 6 beats (shown by the green D4 box at the 6th beat around 1.2 seconds). After the attack, discriminator D4 remained False.

demonstrates the trade-off between effectiveness and stealthiness: in order to increase effectiveness, it is inevitable for the attack to become less stealthy. Having access to the full set of Pareto-optimal attacks provides a key advantage to the attacker, who can use the Pareto front to better guide their decision-making process (e.g., by prioritizing effectiveness over stealthiness). In this work, we apply a widely-used criterion to select the ‘‘best’’ solution from a Pareto front: we select the solution with the shortest ℓ_2-distance from the ideal attack, i.e., the one attaining maximum effectiveness and stealthiness (corresponding to the top-right corner in Figure 7).

3) ATTACK TESTING SUCCESS

Figure 8 shows the average attack success across test EGM samples for all threat models, for attacks rendered using the *square pulse* and *white Gaussian noise* templates respectively. For each test case, the attack signal was generated using the best Pareto-optimal parameters obtained during training. Results are averaged over all heart conditions. Specifically for each test sample, we have four possible models: one trained with data from the same patient (i.e., Insider Attacker), one trained with data from a different patient with the same heart condition (Knowledgeable Attacker), and two with population data from FP and FN conditions (Universal Attacker-FP and Universal Attacker-FN), respectively. Our framework yielded more successful Universal Attacker attacks when the two cases were separated than when all of the data were combined;[3] this relaxation is still realistic, as an adversary might try both types of attacks if nothing about the patient is known.

[3] It is difficult, if not impossible, to find a signal perturbation that introduces FPs for some signals and FNs for others.

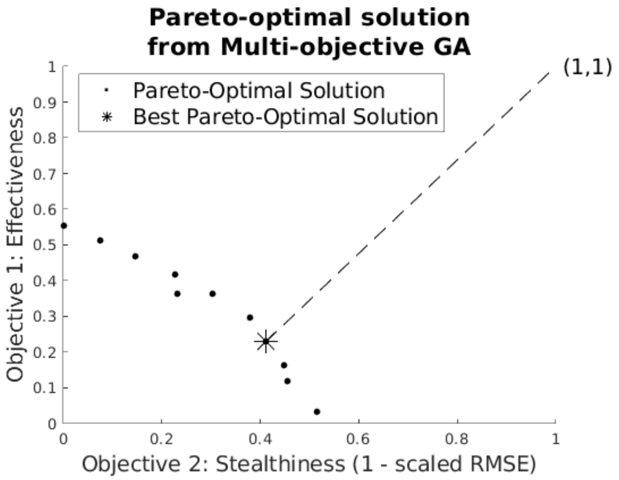

FIGURE 7. **Pareto-optimal solutions for an FN trace. We choose as the "best" solution in the set the one with shortest distance (dashed line) to the ideal point in the objective space, i.e., (1, 1) (100% effective and most stealthy).**

The results of Figure 8 allow us to understand how the probability of attack success is affected by specific limitations on stealthiness. We report average success rates on test data across all the attacks that are *at least* as stealthy as a given threshold. Figure 8 shows how success rates evolve across different percentiles values (0%, 25%, 50%, and 75%) for the stealthiness metric f_S. For example, the average success probability of about 0.3 for the Knowledgeable Attacker model using square pulse attacks at 50% stealthiness is obtained by considering all synthesized Pareto-optimal attacks with stealthiness values in the top 50% of all attacks.

As expected, access to patient-specific information (actual or similar historical data) used by the Insider Attacker and Knowledgeable Attacker models generally led to more successful attacks than the other threat models at each stealthiness threshold. We also observe that the square-pulse attack pattern is consistently more effective than the Gaussian-pulse attack. This may imply that randomized perturbations are not as effective as deterministic attack patterns.

Another finding is that the decay rate of attack success as stealthiness increases is not constant: most models exhibit a greater decline in the success rate in the 50-100% stealthiness range than in the 0-50% range. Our approach was able to find successful attacks under the Universal Attacker threat model as well, although at lower stealthiness levels. We observe that Universal Attacker-FN attacks generally performed worse than their FP counterparts. These attacks indeed require the suppression/delay of multiple beats, meaning that they must be timed precisely to align with ventricular intervals. On the other hand, FP attacks are comparably easier to conduct, as they can introduce peaks at any time point in order to successfully shorten the ventricular period.

Figure 8 reveals some interesting cases where comparisons between threat models are not as expected. For example, Knowledgeable Attacker performed slightly better than Insider Attacker for Gaussian attacks at the 50% stealthiness threshold. While the difference is small, it can be explained by the stealthiness distribution of attacks at the higher percentiles. Figure 9 provides additional insights into the distribution of attacks generated by each model and corroborates this finding. Each point represents attacks on 20 60-second traces for a given heart condition. Insider Attacker attacks are generally much more stealthy than other models. This view also corroborates the expectation that stealthier attacks are less effective within a threat model.

The spread in points in Figure 9 for a given attack type and threat model also illustrates that some heart conditions are more vulnerable to certain types of attacks than others. For example, Knowledgeable Attacker points are clustered more closely together in the Gaussian noise results than in the square-pulse results. This suggests that Gaussian noise attacks induce similar changes in EGM signals across all heart conditions, while the effects of square pulses are more differentiated.

VI. RELATED WORK

Yan et al. [6] consider the general problem of systematizing knowledge of analog attacks against sensor circuitry and corresponding defenses. Their main contribution is a sensor security model that engineers can use to better express analog security properties of circuitry without needing to learn a significantly new notation. Their model introduces transfer functions and a vector of adversarial noise to represent adversarial capabilities at each stage of a sensor's signal-conditioning chain.

Halperin et al. [2] were the first to show that ICDs can be manipulated by wireless attacks, using off-the-shelf hardware. They reverse-engineered an ICD's communication protocol and used software radios to deliver a range of attacks, including reprogramming attacks (that alter the device's parameters to induce inappropriate therapy) and battery depletion attacks. Kune et al. [4] furthered this work by measuring the vulnerability of a number of recently-approved ICDs to various signal-injection attacks across a range of distances and power levels. They found that the devices were susceptible to baseband and amplitude-modulated, low-power attacks. They also introduced possible defenses, including shielding and detection via a signal-corruption metric.

More recent work has looked at synthesizing stealthy attacks. Paoletti et al. [5] presented a formal approach to formulating stealthy reprogramming attacks on ICDs. The authors show that slight deviations in programmed parameters can result in improper therapy decisions. These deviations can be systematically crafted, requiring only knowledge of the ICD model and availability of population-level EGM recordings (akin to our knowledgeable attacker threat model).

Related work also involves model-based approaches to medical-device security. This includes attacking electrocardiogram (ECG)-based biometric authentication [24], a study of the adversarial robustness of deep-learning systems for

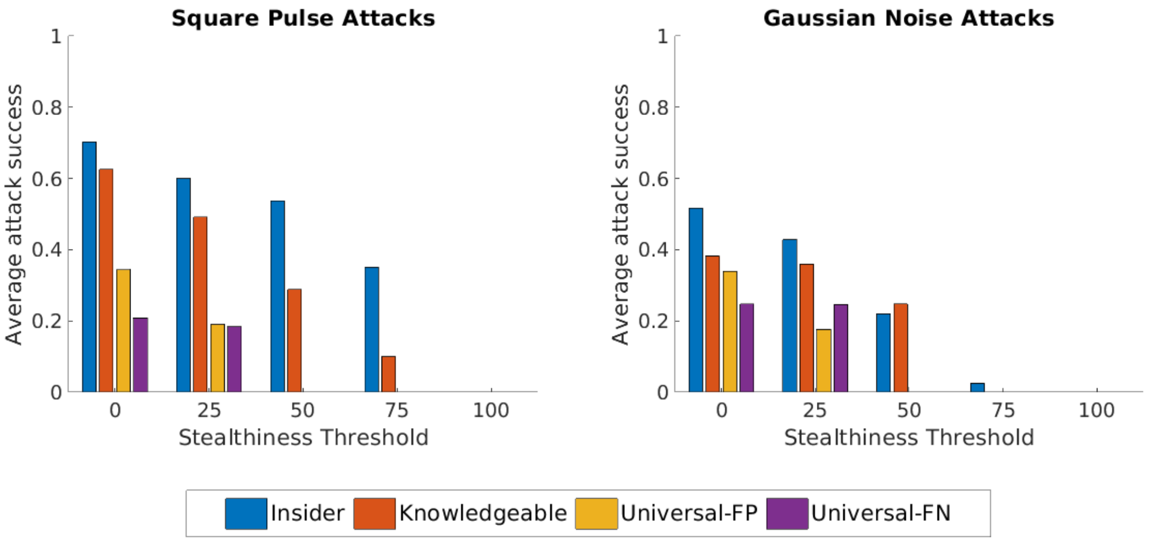

FIGURE 8. Average attack success rate vs. stealthiness level across test sets. Each ($x_{stealth}$, $y_{success}$) point represents the average success rate of attacks that were *at least* as stealthy as $x_{stealth}$. The $x_{stealth}$ cutoffs are discretized into four percentiles: 0%, 25%, 50%, and 75%. For instance, the leftmost bars represent the average success rate of our Pareto-optimal attacks (selected from the optimal front as per Figure 7), over all stealthiness levels. The second group of bars represents the average success rate for attacks ranked among the top-75% in terms of stealthiness.

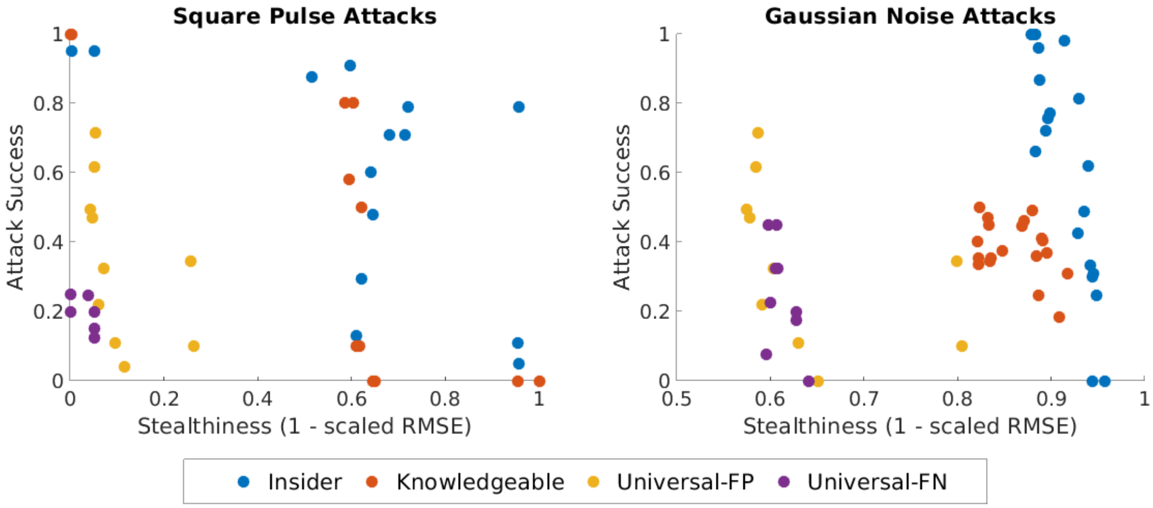

FIGURE 9. Average attack success rate vs. stealthiness level for Pareto-optimal attacks of each attack type. The average success rate of all points of a given threat model and attack type is shown as the first group of bars in Figure 8.

ECG-based arrhythmia detection [25], and attacks on CT scans synthesized using deep generative models [26].

To the best of our knowledge, no existing work has looked at automated methods to derive stealthy and effective signal-injection attacks on cardiac devices: those aimed at inducing incorrect therapy decisions through minimal perturbations of the target cardiac signals.

VII. DISCUSSION

Our results highlight the scope and success rate of stealthy attacks against ICDs under different knowledge levels about potential victims. While Kune et al. showed the feasibility of attacks that completely dominate the true EGM, we have demonstrated how a range of low-power, systematically-devised attacks, centered around pulsed (noisy) signals, can interfere with ICD behavior, while keeping the legitimate signal still highly discernible (i.e., with high Signal-to-Interference ratios, as defined in [4], Eq. 5). Our work is the first to evaluate a range of signal-injection attacks over a large dataset of EGM data. Our results show that attacks can be devised systematically depending on the information available about the patient, and without the need to overpower the legitimate EGM.

Another contribution of our work w.r.t. Kune et al. [4] is that we show that even a very slight modification to a legitimate EGM—short pulses of noise that leave the legitimate waveform discernible to observers —could cause the ICD discrimination algorithm to behave incorrectly. In other words, we expose vulnerabilities in the sensing and discrimination algorithms themselves, rather than in the unshielded electromagnetic components of the device. Our work suggests further investigation into the brittleness of ICD algorithms w.r.t. small amounts of noise and what steps could be taken to improve the robustness of the system.

It is also important to remember that the results we have presented represent the success of the InjectICD framework in suppressing or inducing a shock therapy in 60-second windows. If the attacker were to run longer attacks or multiple attacks at the same time, the likelihood of success could compound over time and approach 1.

A. POSSIBLE DEFENSES

Kune et al. [4] proposed both analog and digital defenses against arbitrary signal injections, which would also defend against the targeted signal-injection attacks introduced in this paper. On the analog side, their proposed defenses include electromagnetic shielding and the application of a conducting material to shield the device from radiation. They showed that even an imperfect ICD shield raised the power requirements for a successful attack. Other hardware-based defenses include incorporating low-pass filters and differential circuits to compare voltage readings against a reference. These are all methods that are well-known and commonly used in analog electronics, but are still absent from most ICDs.

Kune et al. [4] also proposed a number of digital defenses that could prevent or detect the type of attack proposed in this paper. For one, they propose tracking a metric of signal contamination, related to the power of the sensed waveform, to detect anomalies. They moreover developed an adaptive filtering mechanism that consists of a series of analog shields and filters, along with statistical methods to verify the accuracy of signals after digital conversion. These methods would generally prevent the InjectICD low-power injections from succeeding as well.

However, while these defenses were published almost a decade ago, most manufacturers have still not implemented them. The attacks generated by InjectICD still pose a real threat to existing ICDs, particularly to those that are already implanted in patients. Our work demonstrates the scope of possible attacks and raises the urgency for the adoption of the aforementioned defenses, especially since we introduce attacks that can evade visual detection.

VIII. CONCLUSION

In this paper, we introduced the InjectICD framework for synthesizing stealthy and effective signal-injection attacks on ICDs. By formulating the problem as one of multi-objective optimization, our approach can identify parameters for the injected-signal templates that are Pareto-optimal w.r.t. such an effectiveness-stealthiness tradeoff. We considered three realistic threat models, assuming adversaries with tiered access to a victim patient's historical EGMs and heart condition.

We evaluated InjectICD on EGM traces corresponding to 19 heart conditions. Our results show that InjectICD is successful in generating attacks for all heart conditions under the various threat models. Our results also highlight the urgent need for medical device manufacturers to harden their devices against signal-injection attacks.

For future work, we plan to evaluate InjectICD on other ICD models and with more complex attack signals (parameterized by more variables). We also intend to investigate attacks that succeed across longer time frames, such as False Negative attacks that suppress therapy for longer durations. We suspect that these types of attacks will require a sequence of different signal-injection templates. The duration and ordering of multiple templates could also be framed as an optimization problem. Finally, we plan to experimentally validate the InjectICD framework by evaluating our attack strategies *ex vivo* on actual ICD devices.

REFERENCES

[1] G. K. K. Ammannaya, "Implantable cardioverter defibrillators—The past, present and future," *Arch. Med. Sci.-Atherosclerotic Diseases*, vol. 11, pp. e163–e170, Jul. 2020.

[2] D. Halperin, T. S. Heydt-Benjamin, B. Ransford, S. S. Clark, B. Defend, W. Morgan, K. Fu, T. Kohno, and W. H. Maisel, "Pacemakers and implantable cardiac defibrillators: Software radio attacks and zero-power defenses," in *Proc. IEEE Symp. Secur. Privacy (SP)*, Oakland, CA, USA, May 2008, pp. 129–142.

[3] A. Hern. (Aug. 2017). *Hacking Risk Leads to Recall of 500,000 Pacemakers Due to Patient Death Fears*. The Guardian. [Online]. Available: http://www.theguardian.com/technology/2017/aug/31/hacking-risk-recall-pacemakers-patient-death-fears-fda-firmware-update

[4] D. F. Kune, J. Backes, S. S. Clark, D. Kramer, M. Reynolds, K. Fu, Y. Kim, and W. Xu, "Ghost talk: Mitigating EMI signal injection attacks against analog sensors," in *Proc. IEEE Symp. Secur. Privacy*, May 2013, pp. 145–159.

[5] N. Paoletti, Z. Jiang, M. A. Islam, H. Abbas, R. Mangharam, S. Lin, Z. Gruber, and S. A. Smolka, "Synthesizing stealthy reprogramming attacks on cardiac devices," in *Proc. 10th ACM/IEEE Int. Conf. Cyber-Phys. Syst.*, Apr. 2019, pp. 13–22.

[6] C. Yan, H. Shin, C. Bolton, W. Xu, Y. Kim, and K. Fu, "SoK: A minimalist approach to formalizing analog sensor security," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2020, pp. 233–248.

[7] I. Martinovic, D. Davies, M. Frank, D. Perito, T. Ros, and D. Song, "On the feasibility of side-channel attacks with brain-computer interfaces," in *Proc. 21st USENIX Secur. Symp. (USENIX Secur.)*, Bellevue, WA, Aug. 2012, pp. 143–158.

[8] A. Ghani, P. P. H. M. Delnoy, A. R. Ramdat Misier, J. J. J. Smit, A. Adiyaman, J. P. Ottervanger, and A. Elvan, "Incidence of lead dislodgement, malfunction and perforation during the first year following device implantation," *Netherlands Heart J.*, vol. 22, no. 6, pp. 286–291, Jun. 2014.

[9] A. J. Moss, C. Schuger, C. A. Beck, M. W. Brown, D. S. Cannom, J. P. Daubert, N. A. M. Estes, H. Greenberg, W. J. Hall, D. T. Huang, J. Kautzner, H. Klein, S. McNitt, B. Olshansky, M. Shoda, D. Wilber, and W. Zareba, "Reduction in inappropriate therapy and mortality through ICD programming," *New England J. Med.*, vol. 367, no. 24, pp. 2275–2283, Dec. 2012.

[10] J. E. Poole, G. W. Johnson, A. S. Hellkamp, J. Anderson, D. J. Callans, M. H. Raitt, R. K. Reddy, F. E. Marchlinski, R. Yee, T. Guarnieri, M. Talajic, D. J. Wilber, D. P. Fishbein, D. L. Packer, D. B. Mark, K. L. Lee, and G. H. Bardy, "Prognostic importance of defibrillator shocks in patients with heart failure," *New England J. Med.*, vol. 359, pp. 1009–1017, Sep. 2008.

[11] S. F. Sears, J. D. Hauf, K. Kirian, G. Hazelton, and J. B. Conti, "Posttraumatic stress and the implantable cardioverter-defibrillator patient," *Circulat., Arrhythmia Electrophysiol.*, vol. 4, no. 2, pp. 242–250, Apr. 2011.

[12] M. Madhavan and P. A. Friedman, ‘‘Optimal programming of implantable cardiac-defibrillators,’’ *Circulation*, vol. 128, no. 6, pp. 659–672, Aug. 2013.

[13] R. T. Borne, P. D. Varosy, and F. A. Masoudi, ‘‘Implantable cardioverter-defibrillator shocks: Epidemiology, outcomes, and therapeutic approaches,’’ *JAMA Internal Med.*, vol. 173, pp. 859–865, May 2013.

[14] P. S. Chan, H. M. Krumholz, G. Nichol, and B. K. Nallamothu, ‘‘Delayed time to defibrillation after in-hospital cardiac arrest,’’ *New England J. Med.*, vol. 358, no. 1, pp. 9–17, Jan. 2008.

[15] J. S. Seybold, *Introduction to RF Propagation*. Hoboken, NJ, USA: Wiley, 2005.

[16] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, ‘‘PhysioBank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals,’’ *Circulation*, vol. 101, no. 23, pp. e215–e220, Jun. 2000. [Online]. Available: http://circ.ahajournals.org/content/101/23/e215.full, doi: 10.1161/01.CIR.101.23.e215.

[17] A. W. Mathews. (Feb. 2015). *Anthem: Hacked Database Included 78.8 Million People*. [Online]. Available: https://www.wsj.com/articles/anthem-hacked-database-included-78-8-million-people-1424807364

[18] *Boston Scientific Reference Guide Implantable Cardioverter Defibrillator*. Accessed: Oct. 24, 2022. [Online]. Available: https://www.bostonscientific.com/content/dam/Manuals/us/current-rev-en/359407-004_multi_RG_en-USA_S.pdf

[19] *Boston Scientific Discrimination Cardiocases*. Accessed: Oct. 24, 2022. [Online]. Available: https://www.cardiocases.com/en/pacingdefibrillation/specificities/discrimination/boston-scientific/boston-scientific-discrimination

[20] Z. Jiang, H. Abbas, K. J. Jang, M. Beccani, J. Liang, S. Dixit, and R. Mangharam, ‘‘In-silico pre-clinical trials for implantable cardioverter defibrillators,’’ in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 169–172.

[21] (2018). *Ann Arbor Electrogram Libraries*. [Online]. Available: http://electrogram.com/

[22] R. D. Berger, D. R. Lerew, J. M. Smith, C. Pulling, and M. R. Gold, ‘‘The rhythm ID going head to head trial (RIGHT): Design of a randomized trial comparing competitive rhythm discrimination algorithms in implantable cardioverter defibrillators,’’ *J. Cardiovascular Electrophysiol.*, vol. 17, no. 7, pp. 749–753, Jul. 2006.

[23] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms* (Wiley-Interscience Series in Systems and Optimization), 1st ed. New York, NY, USA: Wiley, 2001.

[24] S. Eberz, N. Paoletti, M. Roeschlin, A. Patani, M. Kwiatkowska, and I. Martinovic, ‘‘Broken hearted: How to attack ECG biometrics,’’ in *Proc. Netw. Distrib. Syst. Secur. Symp.*, 2017, pp. 1–15.

[25] X. Han, Y. Hu, L. Foschini, L. Chinitz, L. Jankelson, and R. Ranganath, ‘‘Deep learning models for electrocardiograms are susceptible to adversarial attack,’’ *Nature Med.*, vol. 26, no. 3, pp. 360–363, Mar. 2020.

[26] Y. Mirsky, T. Mahler, I. Shelef, and Y. Elovici, ‘‘CT-GAN: Malicious tampering of 3D medical imagery using deep learning,’’ in *Proc. 28th USENIX Secur. Symp. (USENIX Secur.)*, 2019, pp. 461–478.

VEENA KRISH received the bachelor’s and master’s degrees in bioengineering from the University of Pennsylvania. She is currently pursuing the Ph.D. degree with the Department of Computer Science, Stony Brook University. Her research interests include security aspects of emerging medical technologies and machine learning components.

NICOLA PAOLETTI received the Ph.D. degree in information sciences and complex systems from the Università di Camerino, Italy. He was a Lecturer at the Department of Computer Science, Royal Holloway, University of London, U.K., and a Postdoctoral Researcher at Stony Brook University, USA, and the University of Oxford, U.K. He is a Senior Lecturer with the Department of Informatics, King’s College London, U.K. His research interests include safety and security assurance of cyber-physical systems (CPSs), with an emphasis on biomedical applications, and formal analysis methods (verification, control, and synthesis) to design CPSs that are provably correct, including robustness and uncertainty analysis of systems with machine-learning components in the loop.

SCOTT A. SMOLKA is a SUNY Distinguished Professor of computer science with Stony Brook University. He is the Lead PI for the multi-institutional NSF CPS Frontiers project on CyberCardia: Compositional, Approximate, and Quantitative Reasoning for Medical Cyber-Physical Systems. He has published more than 200 publications, generating more than 10,000 citations. He is a PI/Co-PI on grants totaling more than $30 million. His research interests include model checking, runtime verification, and the modeling and analysis of cyber-physical systems, including cardiac tissue, medical devices, neural circuits, and multi-agent flight formation. He is perhaps best known for the algorithm he and Paris Kanellakis invented for deciding bisimulation. He is an ACM Fellow and a fellow of the European Association for Theoretical Computer Science. He was a co-recipient of the 2021 ACM Edsger J. Dijkstra Award.

AMIR RAHMATI received the Ph.D. degree in computer science and engineering from the University of Michigan, in 2017. He is an Assistant Professor with the Department of Computer Science, Stony Brook University, where he leads the Ethos Security and Privacy Laboratory. His works involve designing, building, and evaluating systems that tackle security challenges in these domains. His research has received frequent attention from media outlets, including MIT Technology Review, Washington Post, and Bloomberg. His work on the security of autonomous driving systems is currently on display with the London Science Museum. His research interests include security of emerging technologies, machine learning, and the Internet of Things.