FRAMEWORK FOR SYNTHESIZING ATTACKS ON ICDs

Veena Krish, Nicola Paoletti, Scott Smolka, Amir Rahmati
Ethos Lab, Advised By: Prof Amir Rahmati

Stony Brook University
Computer Science

INTRODUCTION

Problem: How can we evaluate the robustness of algorithms used in Implantable Cardioverter-Defibrilators (ICDs): medical devices that monitor heart signals (EGMs) and administer therapy?

Previous Work: has shown how heavy-handed electromagnetic interference [1] and slight reprogramming attacks [2] could each thwart therapy; yet, the range of possible damage from adversaries with various capabilities is not well-studied.

Our goal: Devise stealthy attacks against one real-time system (Rhythm ID used in Boston Scientific ICDs), as part of a broader effort to define the limits of attacks on medical cyber-physical systems.

KEY FINDINGS

> Therapy discrimination algorithms that operate on simple, single-beat features expose potential for short-lived (single-beat) attacks.

> Deconstruction of the *Rhythm ID* algorithm into features and states assigned at each heartbeat can guide the search for malicious inputs

ATTACK SYNTHESIS

- Interference modeled as added white gaussian noise (each of 3 heart signals)
- An adversary that has access to historical data can systematically identify features to use as a surrogate loss function *(Algm 1)*
- Stealthy attack synthesis (*Algm 2*) formulated as a **multi-objective optimization**: we aim to manipulate inputs that
  1. maximize selected feature *(from Algm 1)*
  2. minimize detectability (minimize signal-to-noise, RMSE)

Fig 1. Shows two examples, each modified using attack parameters resulting from the following procedures

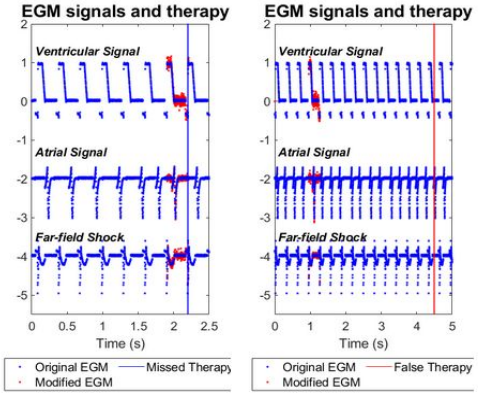
Algorithm 1 Determine Surrogate Loss

```
for each target beat, each historical trace for patient do
    Identity most common internal state changes required for therapy
    L_surrogate ← first feature that causes specific state change
end for
```

Algorithm 2 Optimize attack parameters given target feature

```
for each historical trace do
    L_surrogate = DetermineSurrogateLoss()
    attack_parameters ← SNR, duration of attack, position of start
    Minimize (1) surrogate loss feature, (2) stealthiness wrt. attack parameters
    Evaluate success of attack with solved parameters
end for
```

FIG 1: TWO EXAMPLES OF SHORT-LIVED ATTACKS FOUND VIA MULTI-OBJECTIVE OPTIMIZATION



NEXT STEPS

This work is a key component of a larger framework to strengthen devices; we plan to also investigate:

- Limited threat models (access to data from wearables)
- Extent of victim personalization
- More complex models (deep learning, systems of model control + machine learning)

[1] 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," inSecurity and Privacy (SP). IEEE, 2013, pp. 145–159.
[2] 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 10th International Conference on Cyber-Physical Systems. IEEE, 2019.